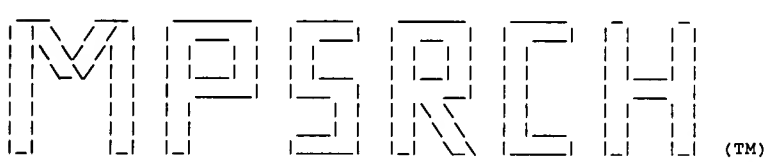
****



****

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Jun 29 19:59:11 1997; MasPar time 1159.38 Seconds
903.507 Million cell updates/sec

Tabular output not generated.

Title: >US-08-663-618A-1
Description: (1-1636) from US08663618A.seq
Perfect Score: 1636
N.A. Sequence: 1 CATGGTGCGGTCTGTGGCCT.....CCCTCTTCCCAAAAAAAAAA 1636
Comp: GTACCACGCCAGACACCGGA.....GGGAGAAGGGTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 886179 seqs, 320143548 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: EST-STS
1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98
99:EST99

Database: EST-STS-TWO
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134
135:EST135 136:EST136 137:EST137 138:EST138 139:EST139
140:EST140 141:EST141 142:EST142 143:EST143 144:EST144
145:EST145 146:EST146 147:EST147 148:EST148 149:EST149
150:EST150 151:EST151 152:EST152 153:EST153 154:EST154
155:EST155 156:EST156 157:EST157 158:EST158 159:EST159
160:EST160 161:EST161 162:EST162 163:EST163 164:EST164
165:EST165 166:EST166 167:EST167 168:EST168 169:EST169
170:EST170 171:EST171 172:EST172 173:EST173 174:EST174
175:EST175 176:EST176 177:EST177 178:EST178 179:EST179
180:EST180 181:EST181 182:EST182 183:EST183 184:EST184
185:EST185 186:EST186 187:EST187 188:EST188 189:EST189
190:EST190 191:EST191 192:EST192 193:EST193

Database: EST-STS-THREE
194:STS1 195:STS2 196:STS3 197:STS4 198:STS5 199:STS6
200:STS7 201:STS8 202:STS9 203:STS10 204:STS11 205:STS12
206:gnEST1 207:gnEST2 208:gnEST3 209:gnEST4 210:gnEST5
211:gnEST6 212:gnEST7 213:gnEST8 214:gnEST9 215:gnEST10
216:gnEST11 217:gnEST12 218:gnEST13 219:gnEST14
220:gnEST15 221:gnEST16 222:gnSTS 223:enEST1 224:enEST2
225:enEST3 226:enEST4 227:enEST5 228:enEST6 229:enEST7
230:enEST8 231:enEST9 232:enEST10 233:enEST11
234:enEST12 235:enEST13 236:enEST14 237:enEST15
238:enEST16 239:enEST17 240:enEST18 241:enEST19
242:enEST20 243:enEST21 244:enEST22 245:enEST23
246:enEST24 247:enEST25 248:enEST26 249:enEST27
250:enSTS

Statistics: Mean 11.516; Variance 2.174; scale 5.296

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
c	1	348	21.3	428	151	T94579	ye31g05.s1 Homo sapie	0.00e+00
	2	245	15.0	405	151	T94272	ye31g05.r1 Homo sapie	0.00e+00
c	3	218	13.3	241	250	HSG124	human STS SHGC-17364.	0.00e+00
c	4	218	13.3	241	202	G29124	human STS SHGC-17364.	0.00e+00
	5	81	5.0	405	41	H43270	yp05a12.r1 Homo sapie	1.04e-104
	6	78	4.8	541	113	N75653	yv29c01.r1 Homo sapie	8.69e-99
	7	78	4.8	541	207	N75653	yv29c01.r1 Homo sapie	8.69e-99
c	8	73	4.5	533	144	T65854	yc24b03.s1 Homo sapie	5.24e-89
c	9	73	4.5	533	200	G24226	human STS WI-14165.	5.24e-89
	10	71	4.3	311	79	N40107	yw72b09.r1 Homo sapie	3.96e-85
	11	71	4.3	336	169	W51302	ma16f08.r1 Life Tech	3.96e-85
	12	66	4.0	266	35	H24821	yl42f01.r1 Homo sapie	1.59e-75
	13	65	4.0	396	134	R55530	yj79c12.r1 Homo sapie	1.28e-73
	14	60	3.7	306	215	W10705	ma47g07.r1 Soares mou	3.39e-64
	15	60	3.7	306	174	W10705	ma47g07.r1 Soares mou	3.39e-64
	16	60	3.7	421	139	R74158	yj99g05.r1 Homo sapie	3.39e-64
c	17	60	3.7	438	76	N26985	yw72b09.s1 Homo sapie	3.39e-64
	18	60	3.7	586	170	W53184	md19a03.r1 Soares mou	3.39e-64
	19	59	3.6	589	193	W97818	mg03c04.r1 Soares mou	2.48e-62
	20	56	3.4	320	144	T66009	yc24b03.r1 Homo sapie	8.71e-57
	21	49	3.0	299	151	T91693	ye25h05.r1 Homo sapie	3.79e-44
c	22	43	2.6	459	31	H10989	ym07a06.s1 Homo sapie	9.88e-34
	23	41	2.5	461	56	H94102	yw58g11.r1 Homo sapie	2.30e-30
c	24	38	2.3	392	95	N64805	yz31c01.s1 Homo sapie	1.98e-25
	25	36	2.2	335	47	CELK104H2F	C.elegans cDNA clone	3.14e-22
	26	35	2.1	280	98	R03212	pk05a03.s1 Caenorhabd	1.17e-20
	27	35	2.1	360	46	CELK084F3F	C.elegans cDNA clone	1.17e-20
	28	34	2.1	360	47	CELK104C7F	C.elegans cDNA clone	4.17e-19
	29	34	2.1	360	45	CELK051H2F	C.elegans cDNA clone	4.17e-19
	30	34	2.1	429	44	CELK009D7F	C.elegans cDNA clone	4.17e-19
	31	32	2.0	323	16	AA067770	mm24c01.r1 Stratagene	4.51e-16
	32	32	2.0	323	244	MMA67770	mm24c01.r1 Stratagene	4.51e-16
	33	33	2.0	360	45	CELK060G3F	C.elegans cDNA clone	1.41e-17
	34	33	2.0	360	47	CELK109D2F	C.elegans cDNA clone	1.41e-17
	35	31	1.9	379	224	CEC4671	C.elegans cDNA clone	1.36e-14
	36	31	1.9	379	39	C08467	C.elegans cDNA clone	1.36e-14
	37	28	1.7	280	231	HSA62630	zm02h01.r1 Stratagene	2.57e-10
	38	28	1.7	280	14	AA062630	zm02h01.r1 Stratagene	2.57e-10
	39	26	1.6	331	11	AA054927	SWMFCA800SK Brugia ma	1.27e-07
	40	26	1.6	331	223	BMA54927	SWMFCA800SK Brugia ma	1.27e-07
	41	24	1.5	375	185	W18653	mb98b05.r1 Soares mou	4.42e-05
	42	24	1.5	375	218	W18653	mb98b05.r1 Soares mou	4.42e-05
	43	25	1.5	447	200	G22362	human STS WI-14048.	2.48e-06

```
   44     25    1.5    447 109  R40271      yf80f04.s1 Homo sapie  2.48e-06
   45     24    1.5    450 216  W13185      ma81d02.r1 Soares mou  4.42e-05
```

ALIGNMENTS

```
RESULT    1
LOCUS       T94579         428 bp    mRNA            EST       24-MAR-1995
DEFINITION  ye31g05.s1 Homo sapiens cDNA clone 119384 3'.
ACCESSION   T94579
NID         g728067
KEYWORDS    EST.
SOURCE      human clone=119384 library=Stratagene lung (#937210)
            vector=pBluescript SK- host=SOLR cells (kanamycin resistant)
            primer=-21m13 Rsite1=EcoRI Rsite2=XhoI Normal lung tissue from a 72
            year old male. Cloned unidirectionally. Primer: Oligo dT. Average
            insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:
            5'-GAATTCGGCACGAG-3'; 3' adaptor sequence:
            5'-CTCGAGTTTTTTTTTTTTTTTT-3'.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 428)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 241
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1..428
                     /organism="Homo sapiens"
                     /clone="119384"
BASE COUNT      112 a    103 c    124 g     76 t     13 others
ORIGIN

  Query Match           21.3%;  Score 348;  DB 151;  Length 428;
  Best Local Similarity  94.4%;  Pred. No. 0.00e+00;
  Matches   386;  Conservative    0;  Mismatches  18;  Indels   5;  Gaps   5;

Db      4 gaagangggcacaaaccaaagatttattttgcaagtgaaaggggcagcc-aggagaccca 62
          ||||| ||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Cp   1624 GAAGAGGGGCACAAACCAAAGATTTATTTTGCAAGTGAAAGGGGCAGCCCAGGAGACCCA 1565

Db     63 gaaaaaaggaaggcaaggctgagagcagaaagcctggataaaggaagaccacagaaaggc 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1564 GAAAAAAGGAAGGCAAGGCTGAGAGCAGAAAGCCTGGATAAAGGAAGACCACAGAAAGGC 1505

Db    123 ctgcaggagccagattgcggcccccagggaaaacccaggaggcaggctgtagagtgatcc 182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1504 CTGCAGGAGCCAGATTGCGGCCCCCAGGGAAAACCCAGGAGGCAGGCTGTAGAGTGATCC 1445

Db    183 tgggcccagcctcaaagctggacttggaggggctttan-gactcaattccaggtgcagca 241
          |||||||||||||||||||   |||||||||||||  ||||||||||||||||||||||
Cp   1444 TGGGCCCAGCCTCAAAGCTGGGACTGGAGGGGCTTTAGCGACTCAATTCCAGGTGCAGCA 1385

Db    242 tttgcagganttgctgaacaccaggcctntcgggcagcttt-ctgnaacaccngccccgn 300
          |||||||| |||||||||||||||||| |||||||||||| ||| |||| | ||||||
Cp   1384 TTTGCAGGAGTTGCTGAACACCAGGCCTGTCGGGCAGCTTTGCTGGAACAGCCGCCCCGC 1325

Db    301 tgcacaactgtagaagctggaccgttcccgaggnttgggatagagcccatcagcttt-cc 359
          |||||| |||||||||||||||||||||||| |||||||||||||||||||||||| ||
Cp   1324 TGCACAGCTGTAGAAGCTGGACCGTTCCCGAGGATTGGGATAGAGCCCATCAGCTTTGCC 1265

Db    360 ctggcagaacgtttcttgtncagggcttngggcatgcttcaggttcaga 408
          |||||||||||| |||||| |||||||  || |||||| ||||||||||
Cp   1264 CTGGCAGAACGTGTCTTGTCCAGGGCTGGGGCCATGCT-CAGGTTCAGA 1217


RESULT    2
LOCUS       T94272         405 bp    mRNA            EST       24-MAR-1995
DEFINITION  ye31g05.r1 Homo sapiens cDNA clone 119384 5'.
ACCESSION   T94272
NID         g727760
KEYWORDS    EST.
SOURCE      human clone=119384 library=Stratagene lung (#937210)
            vector=pBluescript SK- host=SOLR cells (kanamycin resistant)
            primer=M13RP1 Rsite1=EcoRI Rsite2=XhoI Normal lung tissue from a 72
            year old male. Cloned unidirectionally. Primer: Oligo dT. Average
            insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:
            5'-GAATTCGGCACGAG-3'; 3' adaptor sequence:
            5'-CTCGAGTTTTTTTTTTTTTTTT-3'.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 214
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1..405
                     /organism="Homo sapiens"
                     /clone="119384"
BASE COUNT       64 a    113 c    102 g    115 t     11 others
ORIGIN

  Query Match           15.0%;  Score 245;  DB 151;  Length 405;
  Best Local Similarity  91.6%;  Pred. No. 0.00e+00;
  Matches   371;  Conservative    0;  Mismatches  14;  Indels  20;  Gaps  20;

Db      1 tctgaacctgagcatggccccagccctggacaagacacgttctgccagggcaaagctgat 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1217 TCTGAACCTGAGCATGGCCCCAGCCCTGGACAAGACACGTTCTGCCAGGGCAAAGCTGAT 1276

Db     61 gggctctatcccaatcctcgggaacggtccagcttctacagctgtgcagcggggcg-ctg 119
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||| |
Qy   1277 GGGCTCTATCCCAATCCTCGGGAACGGTCCAGCTTCTACAGCTGTGCAGCGGGGCGGCTG 1336

Db    120 ttccagcaaagctgcccgacag-cctggtgttcagcaactcctgcaaatgctgcacctgg 178
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Qy   1337 TTCCAGCAAAGCTGCCCGACAGGCCTGGTGTTCAGCAACTCCTGCAAATGCTGCACCTGG 1396

Db    179 gaattgagtcgctaaagcccctccagtcccagctttgaggctgggcccagggattcactc 238
          |||||||||||||||||||||||||||||||||||||||||||||||||||| || |||||
Qy   1397 -AATTGAGTCGCTAAAGCCCCTCCAGTCCCAGCTTTGAGGCTGGGCCCAGG-AT-CACTC 1453

Db    239 ttacagccttgccttcctggggttttccctgggggccgcaatctgggntccttgcaggc 298
          | ||||||| |||| |||||| |||||||||||| |||||||||||  |||| ||||||
Qy   1454 T-ACAGCCT-GCCT-CCTGGG-TTTTCCCTGGGGG-CCGCAATCTGG-CTCCT-GCAGGC 1506
```

```
Db    299 ctttttttgtggtnttccttttantccagggtttnttgttcttcagcnttgcnttcctttt 358
          |||| | ||||| || |||||| |||||| ||| | | ||| |||| |||| ||||||||
Qy   1507 CTTTCT-GTGGTCTT-CCTTTA-TCCAGGCTTTCT-GCTCT-CAGCCTTGCCTTCCTTTT 1561

Db    359 ttnttgggtttcctggggttgcccntttcatttngcaaantaaat 403
          || | |||| ||||||| | |||| ||||| || ||||| |||||
Qy   1562 TTCT-GGGTCTCCTGGGCT-GCCCCTTTCACTT-GCAAAATAAAT 1603


RESULT    3
ID   HSG124     standard; DNA; STS; 241 BP.
AC   G29124;
NI   g1592433
DT   04-OCT-1996 (Rel. 49, Created)
DT   05-OCT-1996 (Rel. 49, Last updated, Version 2)
DE   human STS SHGC-17364.
KW   primer; sequence tagged site; STS sequence.
OS   Homo sapiens (human)
OC   Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC   Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
RN   [1]
RP   1-241
RA   Myers R.M.;
RT   ;
RL   Unpublished.
CC   Contact: Richard M. Myers Stanford Human Genome Center (SHGC)
CC   Stanford University School of Medicine Department of Genetics,
CC   M-344, Stanford, CA 94305, USA Tel: 4157259687 Fax: 4157259689
CC   Email: myers@shgc.stanford.edu Primer A: AAAGCCTGGATAAAGGAAGACC
CC   Primer B: GCTGCACCTGGAATTGAGTC STS size: 150 PCR Profile: Initial
CC   incubation: 94 degrees C for 90 seconds Denaturation: 94 degrees C
CC   for 15 seconds Annealing: 62 degrees C for 23 seconds
CC   Polymerization: 72 degrees C for 30 seconds PCR Cycles: 30 Thermal
CC   Cycler: Perkin Elmer 9600 Protocol: Template: 25 ng Primer: each 1
CC   uM dNTPs: each 200 uM Taq Polymerase: 0.05 units/ul Total Vol: 10
CC   ul Buffer: MgCl2: 2.5 mM KCl: 50 mM Tris-HCl: 20 mM pH: 8.3
CC   Prepared with primer pairs derived from T94579--Merck/UniEST.
FH   Key             Location/Qualifiers
FH
FT   source          1..241
FT                   /organism="Homo sapiens"
FT                   /map="1"
FT   STS             91..240
FT   primer_bind     91..112
FT   primer_bind     complement(221..240)
SQ   Sequence 241 BP; 78 A; 53 C; 77 G; 30 T; 3 other;

  Query Match           13.3%;  Score 218;  DB 250;  Length 241;
  Best Local Similarity  97.1%;  Pred. No. 0.00e+00;
  Matches   233;  Conservative    0;  Mismatches   5;  Indels   2;  Gaps   2;

Db      4 gaagangggcacaaaccaaagatttattttgcaagtgaaaggggcagcc-aggagaccca 62
          ||||| ||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Cp   1624 GAAGAGGGGCACAAACCAAAGATTTATTTTGCAAGTGAAAGGGGCAGCCCAGGAGACCCA 1565

Db     63 gaaaaaaggaaggcaaggctgagagcagaaagcctggataaaggaagaccacagaaaggc 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1564 GAAAAAAGGAAGGCAAGGCTGAGAGCAGAAAGCCTGGATAAAGGAAGACCACAGAAAGGC 1505

Db    123 ctgcaggagccagattgcggcccccagggaaaacccaggaggcaggctgtagagtgatcc 182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1504 CTGCAGGAGCCAGATTGCGGCCCCCAGGGAAAACCCAGGAGGCAGGCTGTAGAGTGATCC 1445

Db    183 tgggcccagcctcaaagctggacttggaggggctttan-gactcaattccaggtgcagca 241
          |||||||||||||||||||  |||||||||||||  ||||||||||||||||||||||
Cp   1444 TGGGCCCAGCCTCAAAGCTGGGACTGGAGGGGCTTTAGCGACTCAATTCCAGGTGCAGCA 1385


RESULT    4
LOCUS       G29124         241 bp    DNA             STS       04-OCT-1996
DEFINITION  human STS SHGC-17364.
ACCESSION   G29124
NID         g1592433
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 241)
  AUTHORS   Myers,R.M.
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157259687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu

            Primer A: AAAGCCTGGATAAAGGAAGACC
            Primer B: GCTGCACCTGGAATTGAGTC
            STS size: 150
            PCR Profile:
                    Initial incubation: 94 degrees C for 90 seconds

                    Denaturation:       94 degrees C for 15 seconds
                    Annealing:          62 degrees C for 23 seconds
                    Polymerization:     72 degrees C for 30 seconds
                    PCR Cycles:         30
                    Thermal Cycler:     Perkin Elmer 9600
            Protocol:
                    Template:       25 ng
                    Primer:         each 1 uM
                    dNTPs:          each 200 uM
                    Taq Polymerase: 0.05 units/ul
                    Total Vol:      10 ul

            Buffer:
                    MgCl2:          2.5 mM
                    KCl:            50  mM
                    Tris-HCl:       20  mM
                    pH:             8.3

            Prepared with primer pairs derived from T94579--Merck/UniEST.
FEATURES             Location/Qualifiers
     source          1..241
                     /organism="Homo sapiens"
                     /map="1"
     STS             91..240
     primer_bind     91..112
     primer_bind     complement(221..240)
BASE COUNT       78 a     53 c     77 g     30 t      3 others
ORIGIN

  Query Match           13.3%;  Score 218;  DB 202;  Length 241;
  Best Local Similarity  97.1%;  Pred. No. 0.00e+00;
  Matches   233;  Conservative    0;  Mismatches   5;  Indels   2;  Gaps   2;

Db      4 gaagangggcacaaaccaaagatttattttgcaagtgaaaggggcagcc-aggagaccca 62
          ||||| |||||||||||||||||||||||||||||||||||||||||| |||||||||||
Cp   1624 GAAGAGGGGCACAAACCAAAGATTTATTTTGCAAGTGAAAGGGGCAGCCCAGGAGACCCA 1565

Db     63 gaaaaaaggaaggcaaggctgagagcagaaagcctggataaaggaagaccacagaaaggc 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1564 GAAAAAAGGAAGGCAAGGCTGAGAGCAGAAAGCCTGGATAAAGGAAGACCACAGAAAGGC 1505

Db    123 ctgcaggagccagattgcggcccccagggaaaacccaggaggcaggctgtagagtgatcc 182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1504 CTGCAGGAGCCAGATTGCGGCCCCCAGGGAAAACCCAGGAGGCAGGCTGTAGAGTGATCC 1445

Db    183 tgggcccagcctcaaagctggacttggaggggctttan-gactcaattccaggtgcagca 241
          |||||||||||||||||||  |||||||||||||  ||||||||||||||||||||||
Cp   1444 TGGGCCCAGCCTCAAAGCTGGGACTGGAGGGGCTTTAGCGACTCAATTCCAGGTGCAGCA 1385
```

```
RESULT    5
LOCUS       H43270        405 bp    mRNA            EST       31-JUL-1995
DEFINITION  yp05a12.r1 Homo sapiens cDNA clone 186526 5' similar to gb:M80927
            CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   H43270
NID         g919322
KEYWORDS    EST.
SOURCE      human clone=186526 library=Soares breast 3NbHBst vector=pT7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=M13RPl Rsite1=Not I Rsite2=Eco RI Adult human.
            1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
            digested with Not I and cloned into the Not I and Eco RI sites of a
            modified pT7T3 vector (Pharmacia). Library went through one round
            of normalization to a Cot = 20. Library constructed by Bento Soares
            and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 323
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1..405
                     /organism="Homo sapiens"
                     /clone="186526"
BASE COUNT      106 a    112 c     95 g     86 t      6 others
ORIGIN

  Query Match            5.0%;  Score 81;  DB 41;  Length 405;
  Best Local Similarity  63.3%;  Pred. No. 1.04e-104;
  Matches   197;  Conservative    0;  Mismatches 113;  Indels   1;  Gaps   1;

Db      1 gctgcttcccagatgcccttgaccgcttccnctgtacccacatcatctacagctttgcca 60
          ||| | | ||  | | | | |||| |  ||  || |||||| ||||||||  ||| ||
Qy    120 GCTTCCTGCCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTG 179

Db     61 atataagcaacgatcacatcgacacctgggagtggaatgatgtgacgctctacggcatgc 120
            || | |||| | ||  |   ||||  ||||||||||| | ||| ||||||     |
Qy    180 GCATGACCAACCACCAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGT 239

Db    121 tcaacacactcaagaacaggaaccccaacctgaagactctcttgtctgtgggaggatgga 180
          ||||    || ||||| | ||| ||||| |||||||| || ||  |  | ||||| ||||
Qy    240 TCAATGGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGA 299

Db    181 ncnttgggtctcaaagattttccaagatagcctccaacacccagagtcgccggactttca 240
             | ||  |||| |  ||  |  | || |   |||   || | |  || | ||| ||
Qy    300 ATTTCGGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTG 359

Db    241 tcaagtcagtaccgccatttctgcgcacccatggctttgatggggctggaccttgcctgg 300
          |||| || |       ||||||||||    |  ||||||| ||  ||  || ||||||| ||||
Qy    360 TCAACTCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGC-CTTGACCTTGACTGG 418

Db    301 gtctaccctgg 311
          |  ||||| ||
Qy    419 GAGTACCCAGG 429


RESULT    6
LOCUS       N75653        541 bp    mRNA            EST       02-APR-1996
DEFINITION  yv29c01.r1 Homo sapiens cDNA clone 244128 5' similar to gb:M80927
            CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   N75653
NID         g1238231
KEYWORDS    EST.
SOURCE      human clone=244128 primer=reverse ET library=Soares fetal liver
            spleen 1NFLS vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Pac I Rsite2=Eco RI Liver
            and spleen from a 20 week-post conception male fetus. 1st strand
            cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 441
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1..541
                     /organism="Homo sapiens"
                     /clone="244128"
     mRNA            <1..>541
BASE COUNT      149 a    134 c    117 g    136 t      5 others
ORIGIN

  Query Match            4.8%;  Score 78;  DB 113;  Length 541;
  Best Local Similarity  66.7%;  Pred. No. 8.69e-99;
  Matches   166;  Conservative    0;  Mismatches  82;  Indels   1;  Gaps   1;

Db     27 ctgggcaggtgtagtggtcttgctgcttctccagggaggatctgcctacaaactggtttg 86
          |||||||||| |  ||||| ||||| |  |||   | || |||||    |||||||| ||
Qy     19 CTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTG 78

Db     87 ctactttaccaactggtcccaggaccggcagaaccaggaaaa-ttcacccctgagaatat 145
          |||||| ||||||||| ||||| || | |||    |||     |||   ||  || |  |
Qy     79 CTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTT 138

Db    146 tgacccctttcctatgctctcatctcatctattcattcgccagcatcgaaaacaacaaggt 205
          ||||||  ||| ||| | || ||||||||  | |||||  ||||    ||| || || |
Qy    139 GGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCT 198

Db    206 tatcatcaaggacaagagtgaagtgatgctctaccagaccatcaacagtctcaaaaccaa 265
```

```
          | || ||  ||   || ||| | ||  |||||||||    ||||  | || || |  |
Qy    199 GAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGAT 258

Db    266 gaatcccaa 274
          |||||||||
Qy    259 GAATCCCAA 267


RESULT    7
LOCUS       N75653         541 bp    mRNA            EST       02-APR-1996
DEFINITION  yv29c01.r1 Homo sapiens cDNA clone 244128 5' similar to gb:M80927
            CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   N75653
NID         g1238231
KEYWORDS    EST.
SOURCE      human clone=244128 primer=reverse ET library=Soares fetal liver
            spleen 1NFLS vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Pac I Rsite2=Eco RI Liver
            and spleen from a 20 week-post conception male fetus. 1st strand
            cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 441
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.

            NCBI gi: 1238231
FEATURES              Location/Qualifiers
     source           1..541
                      /organism="Homo sapiens"
                      /clone="244128"
                      /note="human"
     mRNA             <1..>541
BASE COUNT      149 a    134 c    117 g    136 t      5 others
ORIGIN

  Query Match           4.8%;  Score 78;  DB 207;  Length 541;
  Best Local Similarity  66.7%;  Pred. No. 8.69e-99;
  Matches   166;  Conservative    0;  Mismatches  82;  Indels   1;  Gaps   1;

Db     27 ctgggcaggtgtagtggtcttgctgcttctccagggaggatctgcctacaaactggtttg 86
          |||||||||| |  ||||| ||||| |  |||   | || |||||    |||||||| ||
Qy     19 CTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTG 78

Db     87 ctactttaccaactggtcccaggaccggcagaaccaggaaaa-ttcacccctgagaatat 145
          |||||| ||||||||| ||||| || | |||    |||     |||   ||  || |  |
Qy     79 CTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTT 138

Db    146 tgacccctttcctatgctctcatctcatctattcattcgccagcatcgaaaacaacaaggt 205
          ||||||  ||| ||| | || |||||||  | |||||  ||||    ||| || || |
Qy    139 GGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCT 198

Db    206 tatcatcaaggacaagagtgaagtgatgctctaccagaccatcaacagtctcaaaaccaa 265
          | || ||  ||   || ||| | ||  |||||||||    ||||  | || || |  |
Qy    199 GAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGAT 258

Db    266 gaatcccaa 274
          |||||||||
Qy    259 GAATCCCAA 267


RESULT    8
LOCUS       T65854         533 bp    mRNA            EST       20-FEB-1995
DEFINITION  yc24b03.s1 Homo sapiens cDNA clone 81581 3' similar to SP:S27879
            S27879 SECRETORY PROTEIN YM-1 PRECURSOR - ;.
ACCESSION   T65854
NID         g674899
KEYWORDS    EST.
SOURCE      human clone=81581 library=Stratagene lung (#937210)
            vector=pBluescript SK- host=SOLR cells (kanamycin resistant)
            primer=-21m13 Rsite1=EcoRI Rsite2=XhoI Normal lung tissue from a 72
            year old male. Cloned unidirectionally. Primer: Oligo dT. Average
            insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:
            5'-GAATTCGGCACGAG-3'; 3' adaptor sequence:
            5'-CTCGAGTTTTTTTTTTTTTTTT-3'.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High qality sequence stops: 307
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1..533
                      /organism="Homo sapiens"
                      /clone="81581"
BASE COUNT      140 a    123 c    136 g    128 t      6 others
ORIGIN

  Query Match           4.5%;  Score 73;  DB 144;  Length 533;
  Best Local Similarity  70.1%;  Pred. No. 5.24e-89;
  Matches   131;  Conservative    0;  Mismatches  55;  Indels   1;  Gaps   1;

Db    250 tggacaagagggtaaggncctggttgcaggatttnccagtgaagtcatccatgtcaata 309
          ||||  || |||||  | ||||||||||||||    || |  |||||||| | ||| |
Cp   1137 TGGATGAGGGGGTATCGGCCCTGGTTGCAGGAGAAGCCGGCAAAGTCATCTAAGTCCAGT 1078

Db    310 gaccagatcatggctcctcccaggtttaaattctttaagaactgaaccttggtctccata 369
          | ||||| |||||| || |||||        | ||| |   |    |||||||| | |
Cp   1077 GCCCAGACCATGGCCCCGCCCAGTCCCTTCTGCTTCAGATAGCTGACCTTGGTTTTGAAG 1018

Db    370 ctcttcacatcatcatagcccacccacttggttccccttgactncgtaggggacctgctg 429
          |||| |||||||||| |||||||||||| ||||  ||  ||    |||||| |||| |||
Cp   1017 CTCTCCACATCATCAAAGCCCACCCACT-GGTTGTCCCGGAAGATGTAGGGCACCTTCTG 959

Db    430 atcctgg 436
```

```
          |||||||
Cp    958 ATCCTGG 952


RESULT    9
LOCUS       G24226         533 bp    DNA             STS       31-MAY-1996
DEFINITION  human STS WI-14165.
ACCESSION   G24226
NID         g1344552
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human STSs derived from sequences in dbEST and the Unigene
            collection.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: AGGGCAACAAGTCAACAGCT
            Primer B: TCTCCCAGAATTCTCATGTGG
            STS size: 132
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Derived from dbEST (genbank accession T65854).
FEATURES              Location/Qualifiers
     source           1..533
                      /organism="Homo sapiens"
                      /map="461.5 cR from top of Chr1 linkage group"
     STS              34..165
     primer_bind      34..53
     primer_bind      complement(145..165)
BASE COUNT      140 a    123 c    136 g    128 t      6 others
ORIGIN

  Query Match             4.5%;  Score 73;  DB 200;  Length 533;
  Best Local Similarity  70.1%;  Pred. No. 5.24e-89;
  Matches   131;  Conservative    0;  Mismatches  55;  Indels   1;  Gaps   1;

Db    250 tggacaagagggtaaggnccctggttgcaggatttnccagtgaagtcatccatgtcaata 309
          ||||  || |||||  | ||||||||||||||    || |  |||||||| | ||| |
Cp   1137 TGGATGAGGGGGTATCGGCCCTGGTTGCAGGAGAAGCCGGCAAAGTCATCTAAGTCCAGT 1078

Db    310 gaccagatcatggctcctcccaggtttaaattctttaagaactgaaccttggtctccata 369
          | ||||| |||||| || |||||       | ||| |   |    |||||||| |  |
```

```
Cp   1077 GCCCAGACCATGGCCCCGCCCAGTCCCTTCTGCTTCAGATAGCTGACCTTGGTTTTGAAG 1018

Db    370 ctcttcacatcatcatagcccacccacttggttccccttgactncgtaggggacctgctg 429
          |||| |||||||||| |||||||||||| ||||  ||  ||    |||||| |||| |||
Cp   1017 CTCTCCACATCATCAAAGCCCACCCACT-GGTTGTCCCGGAAGATGTAGGGCACCTTCTG 959

Db    430 atcctgg 436
          |||||||
Cp    958 ATCCTGG 952


RESULT   10
LOCUS       N40107         311 bp    mRNA            EST       22-JAN-1996
DEFINITION  yw72b09.r1 Homo sapiens cDNA clone 257753 5' similar to gb:M80927
            CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   N40107
NID         g1163652
KEYWORDS    EST.
SOURCE      human clone=257753 primer=T7 library=Soares placenta 8to9weeks
            2NbHP8to9W vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI two
            placentae: one from 8 weeks and another from 9 weeks post
            conception. 1st strand cDNA was primed with a Not I - oligo(dT)
            primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M. Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 268
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1..311
                      /organism="Homo sapiens"
                      /clone="257753"
     mRNA             <1..>311
BASE COUNT       93 a     81 c     68 g     68 t      1 others
ORIGIN

  Query Match             4.3%;  Score 71;  DB 79;  Length 311;
  Best Local Similarity  65.8%;  Pred. No. 3.96e-85;
  Matches   156;  Conservative    0;  Mismatches  80;  Indels   1;  Gaps   1;

Db     69 ctgggcaggtgtagtggtcttgctgcttctccagggaggatctgcctacaaactggtttg 128
          |||||||||| |  ||||| ||||| |  |||   | || |||||    |||||||| ||
Qy     19 CTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTG 78

Db    129 ctactttaccaactggtcccaggaccggcagaaccaggaaaa-ttcacccctgagaatat 187
          |||||| |||||||| ||||| || | |||    |||     |||   ||  || | |
Qy     79 CTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTT 138
```

```
Db    188 tgacccttcctatgctctcatctcatctattcattcgccagcatcgaaaacaacaaggt 247
          ||||||  ||| ||| | || ||||||||  | |||||  ||||    ||| || || |
Qy    139 GGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCT 198

Db    248 tatcatcaaggacaagagtgaagtgatgctctaccagaccatcaacagtctcnaaaa 304
          | || ||  ||   || ||| | ||  |||||||||    ||||  | ||  | ||
Qy    199 GAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAA 255


RESULT   11
LOCUS       W51302        336 bp    mRNA            EST       29-MAY-1996
DEFINITION  ma16f08.r1 Life Tech mouse brain Mus musculus cDNA clone 304743 5'
            similar to gb:X93035 M.musculus mRNA for BRP39 protein (MOUSE);.
ACCESSION   W51302
NID         g1339742
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 336)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:201503
            Seq primer: ETPrimer
            High quality sequence stop: 148.
FEATURES             Location/Qualifiers
     source          1..336
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pCMV-SPORT2; Site_1: SalI; Site_2: NotI;
                     Cloned unidirectionally.  Primer: Oligo dT. pCMV-SPORT2
                     vector."
                     /clone="304743"
                     /clone_lib="Life Tech mouse brain"
                     /dev_stage="adult"
                     /lab_host="DH10B"
     mRNA            <1..>336
BASE COUNT       83 a    100 c     89 g     64 t
ORIGIN

  Query Match             4.3%;  Score 71;  DB 169;  Length 336;
  Best Local Similarity  69.2%;  Pred. No. 3.96e-85;
  Matches   137;  Conservative    0;  Mismatches  60;  Indels   1;  Gaps   1;

Db    140 caggctttgcggtcctgatgctgctccagagctgctctgcgtacaagttggtctgctact 199
          ||||  ||   |||||||  || || |||   |  |||||||    ||  ||||||||||||
Qy     24 CAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTGCTACT 83

Db    200 acaccaactggtcccagtaccgggaaggcaatgggagctgcttcccagacgccctcgacc 259
          ||||||||||| |||||||| |  | ||  | |   ||| | | ||  | | | | ||||
Qy     84 TCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACC 143

Db    260 attccctgtgcacccatatcatctacagctttgcaa-catcagccacaacaagctcagca 318
              ||| ||||||||  ||||||||  ||| ||   ||| | | || || |||| ||||
Qy    144 CCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCTGAGCA 203

Db    319 catccgagtggaatgacg 336
          |  | ||||||||||||||
Qy    204 CCACTGAGTGGAATGACG 221


RESULT   12
LOCUS       H24821        266 bp    mRNA            EST       07-JUL-1995
DEFINITION  yl42f01.r1 Homo sapiens cDNA clone 160921 5' similar to gb:M80927
            CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   H24821
NID         g893720
KEYWORDS    EST.
SOURCE      human clone=160921 library=Soares breast 3NbHBst vector=pT7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=M13RP1 Rsite1=Not I Rsite2=Eco RI Adult human.
            1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
            digested with Not I and cloned into the Not I and Eco RI sites of a
            modified pT7T3 vector (Pharmacia). Library went through one round
            of normalization to a Cot = 20. Library constructed by Bento Soares
            and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 266)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 84
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1..266
                     /organism="Homo sapiens"
                     /clone="160921"
BASE COUNT       58 a     80 c     70 g     51 t      7 others
ORIGIN

  Query Match             4.0%;  Score 66;  DB 35;  Length 266;
  Best Local Similarity  68.3%;  Pred. No. 1.59e-75;
  Matches   127;  Conservative    0;  Mismatches  58;  Indels   1;  Gaps   1;

Db     14 ggaccacagtccatagaatcctcggccagcaggtcccctatgccaccaagggcaaccagt 73
          || ||||    || |||||||  |  ||| |||| |||||   |  |  || ||||||||
Qy    933 GGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCCTACATCTTCCGGGACAACCAGT 992

Db     74 gggtaggatacgacgaccaggaaagcgtcaaaagcaaggtgcagtacctgaaggacaggc 133
          |||| || |  || ||   ||| ||| |||||| ||||||    || |||||| | | |
Qy    993 GGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAGGTCAGCTATCTGAAGCAGAAGG 1052

Db    134 agctgg-cgggcgcatggtatgggccctngacctggatnacttncagggctccttctgcg 192
            |||| ||||  |||||| ||||| || ||| | ||| ||||    |||| || ||||
Qy   1053 GACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGATGACTTTGCCGGCTTCTCCTGCA 1112

Db    193 gccagg 198
          ||||||
Qy   1113 ACCAGG 1118
```

```
RESULT   13
LOCUS       R55530        396 bp    mRNA            EST       22-MAY-1995
DEFINITION  yj79c12.r1 Homo sapiens cDNA clone 154966 5' similar to gb:M80927
            CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   R55530
NID         g824825
KEYWORDS    EST.
SOURCE      human clone=154966 library=Soares breast 2NbHBst vector=pT7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=M13RP1 Rsite1=Not I Rsite2=Eco RI Adult female.
            1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
            digested with Not I and cloned into the Not I and Eco RI sites of a
            modified pT7T3 vector (Pharmacia). Library went through one round
            of normalization to a Cot = 230. Library constructed by Bento
            Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 396)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 284
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1..396
                     /organism="Homo sapiens"
                     /clone="154966"
BASE COUNT       88 a    106 c    109 g     87 t      6 others
ORIGIN

  Query Match             4.0%;  Score 65;  DB 134;  Length 396;
  Best Local Similarity  68.3%;  Pred. No. 1.28e-73;
  Matches   114;  Conservative    0;  Mismatches  53;  Indels   0;  Gaps   0;

Db    125 caggctttgtggtcctggtgctgctccagtgctgctctgcatacaaactggtctgctact 184
          |||| ||  |||||||| || || |||  ||  |||||||    ||||||||||||||||
Qy     24 CAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTGCTACT 83

Db    185 acaccagctggtcccagtaccgggnangcgatgggagctgcttcccagatgcccttgacc 244
          ||||| |||| |||||||| |     | || |   ||| | | ||  | | | | ||||
Qy     84 TCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACC 143

Db    245 gnntcctctgtacccacatcatctacagctttgccaatataagcaac 291
              ||| || |||||| ||||||||  ||| ||    || | ||||
Qy    144 CCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAAC 190


RESULT   14
LOCUS       W10705        306 bp    mRNA            EST       26-APR-1996
DEFINITION  ma47g07.r1 Soares mouse p3NMF19.5 Mus musculus cDNA 5' similar to
            gb:M80927 CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   W10705
NID         g1285179
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 306)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA+ET
            High quality sequence stop: 172.

            NCBI gi: 1285179
FEATURES             Location/Qualifiers
     source          1..306
                     /organism="Mus musculus"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo."
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1..>306
BASE COUNT       78 a     76 c     85 g     67 t
ORIGIN

  Query Match             3.7%;  Score 60;  DB 215;  Length 306;
  Best Local Similarity  72.1%;  Pred. No. 3.39e-64;
  Matches    98;  Conservative    0;  Mismatches  38;  Indels   0;  Gaps   0;

Db     78 agaaggttcccttcgctaccaagggcaaccagtgggtggggtatgaggacaaggagagtg 137
          ||||||| |||| |     |  || ||||||||||||||| | ||| ||   ||||||
Qy    960 AGAAGGTGCCCTACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCT 1019

Db    138 tcaaaaacaaggttgggttcctgaaggagaagaagctggcaggagcatcggtgtgggcac 197
          |||||| ||||||  | |  |||||| |||||   ||||  || ||   ||| |||||||
Qy   1020 TCAAAACCAAGGTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCAC 1079

Db    198 tggatttggatgattt 213
          |||| || ||||| ||
Qy   1080 TGGACTTAGATGACTT 1095


RESULT   15
LOCUS       W10705        306 bp    mRNA            EST       05-SEP-1996
DEFINITION  ma47g07.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 313884 5'
            similar to gb:M80927 CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   W10705
NID         g1285179
KEYWORDS    EST.
```

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 306)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA+ET
            High quality sequence stop: 172.
FEATURES             Location/Qualifiers
     source          1..306
                     /organism="Mus musculus"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
                     /clone="313884"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1..>306
BASE COUNT       78 a     76 c     85 g     67 t
ORIGIN

  Query Match            3.7%;  Score 60;  DB 174;  Length 306;
  Best Local Similarity  72.1%;  Pred. No. 3.39e-64;
  Matches    98;  Conservative    0;  Mismatches  38;  Indels   0;  Gaps   0;

Db     78 agaaggttcccttcgctaccaagggcaaccagtgggtggggtatgaggacaaggagagtg 137
          ||||||| |||| |     |  || ||||||||||||||| | ||| ||   ||||||
Qy    960 AGAAGGTGCCCTACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCT 1019

Db    138 tcaaaaacaaggttgggttcctgaaggagaagaagctggcaggagcatcggtgtgggcac 197
          |||||| ||||||  | |  |||||| |||||   ||||  || ||   ||| |||||||
Qy   1020 TCAAAACCAAGGTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCAC 1079

Db    198 tggatttggatgattt 213
          |||| || ||||| ||
Qy   1080 TGGACTTAGATGACTT 1095


Search completed: Sun Jun 29 20:18:52 1997
Job time : 1181 secs.
```

****

MPSRCH (TM)

****

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Jun 29 19:29:43 1997; MasPar time 1518.90 Seconds
1197.646 Million cell updates/sec

Tabular output not generated.

Title: >US-08-663-618A-1
Description: (1-1636) from US08663618A.seq
Perfect Score: 1636
N.A. Sequence: 1 CATGGTGCGGTCTGTGGCCT.....CCCTCTTCCCAAAAAAAAAA 1636
Comp: GTACCACGCCAGACACCGGA.....GGGAGAAGGGTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 333249 seqs, 555961234 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl-new11
1:BCT 2:FUN 3:GEN 4:HUM1 5:HUM2 6:HUM3 7:INV1 8:INV2
9:INV3 10:INV4 11:INV5 12:INV6 13:INV7 14:ORG 15:MAM
16:VRT 17:PLN 18:PRO1 19:PRO2 20:ROD 21:SYN 22:UNC
23:VIR1 24:VIR2

Database: genbank97
25:BCT1 26:BCT2 27:BCT3 28:BCT4 29:BCT5 30:BCT6 31:BCT7
32:BCT8 33:BCT9 34:GEN1 35:GEN2 36:HTG 37:INV1 38:INV2
39:INV3 40:INV4 41:INV5 42:INV6 43:INV7 44:INV8 45:INV9
46:MAM1 47:MAM2 48:MAM3 49:VRT1 50:VRT2 51:VRT3 52:PAT1
53:PAT2 54:PAT3 55:PAT4 56:PHG 57:PLN1 58:PLN2 59:PLN3
60:PLN4 61:PLN5 62:PLN6 63:PLN7 64:PLN8 65:PLN9 66:PLN10
67:PRI1 68:PRI2 69:PRI3 70:PRI4 71:PRI5 72:PRI6 73:PRI7
74:PRI8 75:PRI9 76:PRI10 77:PRI11 78:PRI12 79:PRI13
80:PRI14 81:ROD1 82:ROD2 83:ROD3 84:ROD4 85:ROD5 86:ROD6
87:ROD7 88:ROD8 89:STR 90:SYN 91:UNA 92:VRL1 93:VRL2
94:VRL3 95:VRL4 96:VRL5 97:VRL6 98:VRL7 99:VRL8 100:VRL9

Database: genbank-new11
101:BCT 102:GEN 103:INV1 104:INV2 105:MAM 106:VRT
107:PHG 108:PLN 109:PRI1 110:PRI2 111:ROD 112:SYN
113:UNA 114:VRL

Database: u-embl48_97
115:part1 116:part2

Statistics: Mean 11.639; Variance 4.516; scale 2.577

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1603	98.0	1633	74	HSU29615	Human chitotriosidase	0.00e+00
2	184	11.2	1741	76	HUMHA3G	Human glycoprotein mR	3.57e-154
3	159	9.7	1418	74	HSU49835	Human YKL-39 precurso	5.02e-128
4	159	9.7	1434	74	HSU58514	Human chitinase precu	5.02e-128
5	157	9.6	1500	74	HSU58515	Human chitinase (HUMT	6.03e-126
6	156	9.5	1733	48	SSGP38KD	S.scrofa 38kDa hepari	6.59e-125
7	156	9.5	1733	48	SSU19900	Sus scrofa 38kDa hepa	6.59e-125
8	156	9.5	1733	15	SSGP38KD	S.scrofa 38kDa hepari	6.59e-125
9	135	8.3	1616	81	MMBRP39	M.musculus mRNA for B	3.23e-103
10	123	7.5	1506	20	MMD757	House mouse; Musculus	5.92e-91
11	123	7.5	1506	81	D87757	House mouse; Musculus	5.92e-91
12	105	6.4	2022	48	SSU43490	Sus scrofa oviductal	8.60e-73
13	97	5.9	2034	47	OAU16719	Ovis aries estrogen d	7.86e-65
14	93	5.7	1596	47	OAU17988	Ovis aries 92kD oestr	6.99e-61
15	93	5.7	2228	67	BABEDOSG	Papio hamadryas anubi	6.99e-61
16	93	5.7	2228	15	PAEDOSG	Papio hamadryas anubi	6.99e-61
17	93	5.7	2525	20	MMOSG	Mouse mRNA for oviduc	6.99e-61
18	93	5.7	2525	85	MUSOGP	Mouse mRNA for oviduc	6.99e-61
19	89	5.4	2198	73	HSU09550	Human oviductal glyco	5.90e-57
20	85	5.2	1994	46	BOV95BOGP	Bovine mRNA for 95 kD	4.70e-53
21	69	4.2	2387	20	MAOSG	Hamster mRNA for ovid	9.34e-38
22	69	4.2	2387	81	HAMOGP	Hamster mRNA for ovid	9.34e-38
23	59	3.6	2276	81	MAU15048	Mesocricetus auratus	1.61e-28
24	52	3.2	2452	43	MSU02270	Manduca sexta chitina	2.91e-22
25	45	2.8	443	72	HSOGP03	Human oviductal glyco	3.14e-16
26	45	2.8	994	83	MMU56900	Mus musculus chitinas	3.14e-16
27	42	2.6	3060	29	KURCHISHI	Kurthia zopfii gene f	9.92e-14
28	41	2.5	506	72	HSOGP07	Human oviductal glyco	6.55e-13
29	39	2.4	2078	44	PEUPJCHI1	Penaeus japonicus adu	2.71e-11
30	40	2.4	2401	25	BACCHIA3	B.circulans chitinase	4.25e-12
c 31	36	2.2	25083	7	CEC04F6	Caenorhabditis elegan	6.27e-09
c 32	36	2.2	25083	38	CELC04F6	Caenorhabditis elegan	6.27e-09
33	34	2.1	1322	7	BPU59689	Brugia pahangi microf	2.12e-07
34	34	2.1	1322	37	BPU59689	Brugia pahangi microf	2.12e-07
35	34	2.1	1322	103	BPU59689	Brugia pahangi microf	2.12e-07
36	34	2.1	1654	37	BRPCHIT	Brugia malayi chitina	2.12e-07
37	33	2.0	490	44	OVU14639	Onchocerca volvulus c	1.19e-06
38	32	2.0	1364	7	BPU59690	Brugia pahangi microf	6.54e-06
39	32	2.0	1364	103	BPU59690	Brugia pahangi microf	6.54e-06
40	32	2.0	1364	37	BPU59690	Brugia pahangi microf	6.54e-06
41	32	2.0	1700	32	SMCHIB	Serratia marcescens c	6.54e-06
42	33	2.0	1836	32	SHCHITB	S.marcescens (BJL200)	1.19e-06
43	31	1.9	1320	54	I14183	Sequence 14 from pate	3.48e-05
44	31	1.9	1364	54	I14182	Sequence 13 from pate	3.48e-05
45	31	1.9	1405	54	I14180	Sequence 9 from paten	3.48e-05

ALIGNMENTS

RESULT 1
LOCUS HSU29615 1633 bp mRNA PRI 04-NOV-1995
DEFINITION Human chitotriosidase precursor mRNA, complete cds.
ACCESSION U29615
NID g1050957
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1633)
AUTHORS Boot,R.G., Renkema,G.H., Strijland,A., van Zonneveld,A.J. and Aerts,J.M.
TITLE Cloning of a cDNA encoding chitotriosidase, a human chitinase produced by macrophages
JOURNAL J. Biol. Chem. 270 (44), 26252-26256 (1995)
MEDLINE 96064695
REFERENCE 2 (bases 1 to 1633)
AUTHORS Boot,R.G.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-1995) Rolf G. Boot, Department of Biochemistry, AMC, E.C. Slater, Institute, University of Amsterdam, Meibergdreef 15, 1105 AZ, Amsterdam, The Netherlands

```
FEATURES             Location/Qualifiers
     source          1..1633
                     /organism="Homo sapiens"
                     /clone="Ch1.1"
                     /clone_lib="macrophage-1 cDNA library of Rolf G. Boot"
                     /cell_type="macrophage"
     sig_peptide     13..75
     CDS             13..1413
                     /codon_start=1
                     /product="chitotriosidase precursor"
                     /db_xref="PID:g1050958"
                     /translation="MVRSVAWAGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPK
                     DLDPSLCTHLIYAFAGMTNHQLSTTEWNDETLYQEFNGLKKMNPKLKTLLAIGGWNFG
                     TQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQ
                     DLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSW
                     EKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASS
                     SDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQKVPYIFRDNQWVGF
                     DDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYPLIQTLRQELSLPYLPS
                     GTPELEVPKPGQPSEPEHGPSPGQDTFCQGKADGLYPNPRERSSFYSCAAGRLFQQSC
                     PTGLVFSNSCKCCTWN"
     mat_peptide     76..1410
                     /note="human chitinase"
                     /product="chitotriosidase"
     polyA_signal    1605..1610
     polyA_site      1633
                     /note="10 A nucleotides"
BASE COUNT      354 a    489 c    443 g    347 t
ORIGIN

  Query Match          98.0%;  Score 1603;  DB 74;  Length 1633;
  Best Local Similarity  99.8%;  Pred. No. 0.00e+00;
  Matches  1621;  Conservative    0;  Mismatches   0;  Indels   3;  Gaps   3;

Db     12 catggtgcggtctgtggcctgggcaggtttcatggtcctgctgatgatcccatggggctc 71
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 CATGGTGCGGTCTGTGGCCTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTC 60

Db     72 tgctgcaaaactggtctgctacttcaccaactgggcccagtacagacagggggaggctcg 131
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 TGCTGCAAAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCG 120

Db    132 cttcctgcccaaggacttggaccccagcctttgcacccacctcatctacgccttcgctgg 191
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 CTTCCTGCCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGG 180

Db    192 catgaccaaccaccagctgagcaccactgagtggaatgacgagactctctaccaggagtt 251
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 CATGACCAACCACCAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTT 240

Db    252 caatggcctgaagaagatgaatcccaagctgaagaccctgttagccatcggaggctggaa 311
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241 CAATGGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAA 300

Db    312 tttcggcactcagaagttcacagatatggtagccacggccaacaaccgtcagacctttgt 371
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301 TTTCGGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGT 360

Db    372 caactcggccatcaggtttctgcgcaaatacagctttgacggccttgaccttgactggga 431
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    361 CAACTCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGA 420

Db    432 gtacccaggaagccaggggagccctgccgtagacaaggagcgcttcacaaccctggtaca 491
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    421 GTACCCAGGAAGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACA 480

Db    492 ggacttggccaatgccttccagcaggaagcccagacctcagggaaggaacgccttcttct 551
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    481 GGACTTGGCCAATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCT 540

Db    552 gagtgcagcggttccagctgggcagacctatgtggatgctggatacgaggtggacaaaat 611
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    541 GAGTGCAGCGGTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAAT 600

Db    612 cgcccagaacctggattttgtcaaccttatggcctacgacttccatggctcttgggagaa 671
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    601 CGCCCAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAA 660

Db    672 ggtcacgggacataacagccccctctacaagaggcaagaagagagtggtgcagcagccag 731
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    661 GGTCACGGGACATAACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAG 720

Db    732 cctcaacgtggatgctgctgtgcaacagtggctgcagaaggggaccccctgccagcaagct 791
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    721 CCTCAACGTGGATGCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCCTGCCAGCAAGCT 780

Db    792 gatccttggcatgcctacctacggacgctccttcacactggcctcctcatcagacaccag 851
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    781 GATCCTTGGCATGCCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAG 840

Db    852 agtgggggccccagccacagggtctggcactccaggccccttcaccaaggaaggagggat 911
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    841 AGTGGGGGCCCCAGCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGAT 900

Db    912 gctggcctactatgaagtctgctcctggaagggggccaccaaacagagaatccaggatca 971
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    901 GCTGGCCTACTATGAAGTCTGCTCCTGGAAGGGGGCCACCAAACAGAGAATCCAGGATCA 960

Db    972 gaaggtgccctacatcttccgggacaaccagtgggtgggctttgatgatgtggagagctt 1031
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    961 GAAGGTGCCCTACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTT 1020

Db   1032 caaaaccaaggtcagctatctgaagcagaagggactgggcggggccatggtctgggcact 1091
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1021 CAAAACCAAGGTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACT 1080

Db   1092 ggacttagatgactttgccggcttctcctgcaaccagggccgataccccctcatccagac 1151
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1081 GGACTTAGATGACTTTGCCGGCTTCTCCTGCAACCAGGGCCGATACCCCCTCATCCAGAC 1140

Db   1152 gctacggcaggaactgagtcttccatacttgccttcaggcaccccagagcttgaagttcc 1211
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1141 GCTACGGCAGGAACTGAGTCTTCCATACTTGCCTTCAGGCACCCCAGAGCTTGAAGTTCC 1200

Db   1212 aaaaccaggtcagccctctgaacctgagcatggccccagccctggacaagacacgttctg 1271
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1201 AAAACCAGGTCAGCCCTCTGAACCTGAGCATGGCCCCAGCCCTGGACAAGACACGTTCTG 1260

Db   1272 ccagggcaaagctgatgggctctatcccaatcctcgggaacggtccagcttctacagctg 1331
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1261 CCAGGGCAAAGCTGATGGGCTCTATCCCAATCCTCGGGAACGGTCCAGCTTCTACAGCTG 1320

Db   1332 tgcagcggggcggctgttccagcaaagctgcccgacaggcctggtgttcagcaactcctg 1391
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1321 TGCAGCGGGGCGGCTGTTCCAGCAAAGCTGCCCGACAGGCCTGGTGTTCAGCAACTCCTG 1380

Db   1392 caaatgctgcacctggaattgagtcg-taaagcccctccagtcc-agctttgaggctggg 1449
        |||||||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Qy   1381 CAAATGCTGCACCTGGAATTGAGTCGCTAAAGCCCCTCCAGTCCCAGCTTTGAGGCTGGG 1440

Db   1450 cccaggatcactctacagcctgcctcctgggttttcc-tgggggccgcaatctggctcct 1508
        |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Qy   1441 CCCAGGATCACTCTACAGCCTGCCTCCTGGGTTTTCCCTGGGGGCCGCAATCTGGCTCCT 1500

Db   1509 gcaggcctttctgtggtcttcctttatccaggctttctgctctcagccttgccttccttt 1568
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1501 GCAGGCCTTTCTGTGGTCTTCCTTTATCCAGGCTTTCTGCTCTCAGCCTTGCCTTCCTTT 1560

Db   1569 tttctgggtctcctgggctgccccctttcacttgcaaaataaatctttggtttgtgccccct 1628
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1561 TTTCTGGGTCTCCTGGGCTGCCCCTTTCACTTGCAAAATAAATCTTTGGTTTGTGCCCCT 1620

Db   1629 cttc 1632
        ||||
Qy   1621 CTTC 1624
```

```
RESULT    2
LOCUS       HUMHA3G      1741 bp    mRNA            PRI       24-JAN-1994
DEFINITION  Human glycoprotein mRNA, complete cds.
ACCESSION   M80927
NID         g348911
KEYWORDS    glycoprotein.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1741)
  AUTHORS   Hakala,B.E., White,C. and Recklies,A.D.
  TITLE     Human cartilage gp-39, a major secretory product of articular
            chondrocytes and synovial cells, is a mammalian member of a
            chitinase protein family
  JOURNAL   J. Biol. Chem. 268 (34), 25803-25810 (1993)
  MEDLINE   94064658
FEATURES             Location/Qualifiers
     source          1..1741
                     /organism="Homo sapiens"
                     /cell_type="chondrocyte"
                     /tissue_type="cartilage"
     5'UTR           1..71
     sig_peptide     72..134
                     /codon_start=1
     CDS             72..1223
                     /note="articular 39kDa glycoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /product="glycoprotein"
                     /db_xref="PID:g348912"
                     /translation="MGVKASQTGFVVLVLLQCCSAYKLVCYYTSWSQYREGDGSCFPD
                     ALDRFLCTHIIYSFANISNDHIDTWEWNDVTLYGMLNTLKNRNPNLKTLLSVGGWNFG
                     SQRFSKIASNTQSRRTFIKSVPPFLRTHGFDGLDLAWLYPGRRDKQHFTTLIKEMKAE
                     FIKEAQPGKKQLLLSAALSAGKVTIDSSYDIAKISQHLDFISIMTYDFHGAWRGTTGH
                     HSPLFRGQEDASPDRFSNTDYAVGYMLRLGAPASKLVMGIPTFGRSFTLASSETGVGA
                     PISGPGIPGRFTKEAGTLAYYEICDFLRGATVHRTLGQQVPYATKGNQWVGYDDQESV
                     KSKVQYLKDRQLAGAMVWALDLDDFQGSFCGQDLRFPLTNAIKDALAAT"
     mat_peptide     135..1220
                     /codon_start=1
                     /product="glycoprotein"
     3'UTR           1224..1741
     polyA_signal    1715..1720
     polyA_site      1741
BASE COUNT      416 a    510 c    447 g    368 t
ORIGIN

  Query Match           11.2%;  Score 184;  DB 76;  Length 1741;
  Best Local Similarity  63.8%;  Pred. No. 3.57e-154;
  Matches   701;  Conservative    0;  Mismatches 373;  Indels  24;  Gaps  10;

Db      94 caggctttgtggtcctggtgctgctccagtgctgctctgcataca-aactggtctgctact 153
           |||| ||  |||||||| || || |||  ||  |||||||    ||||||||||||||||
Qy      24 CAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTGCTACT 83

Db     154 acaccagctggtcccagtaccgggaaggcgatgggagctgcttcccagatgcccttgacc 213
           ||||| |||| |||||||| |  | || || |   ||| | | ||  | | | | ||||
Qy      84 TCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACC 143

Db     214 gcttcctctgtacccacatcatctacagctttgccaatataagcaacgatcacatcgaca 273
           |  ||| || |||||| ||||||||  ||| ||    || | |||| | ||  |   ||
Qy     144 CCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCTGAGCA 203

Db     274 cctgggagtggaatgatgtgacgctctacggcatgctcaacacactcaagaacaggaacc 333
           ||   |||||||||| | ||| ||||||      | ||||    || ||||| | ||| |
Qy     204 CCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGATGAATC 263

Db     334 ccaacctgaagactctcttgtctgtcggaggatggaactttgggtctcaaagattttcca 393
           |||| |||||||| || ||  |  ||||||| ||||| || ||  |||| |  ||  |
Qy     264 CCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCGGCACTCAGAAGTTCACAG 323

Db     394 agatagcctccaacacccagagtcgccggactttcatcaagtcagtaccgccattcctgc 453
           | || |   |||   || | |  || | ||| ||  |||| || |        || ||||

Qy     324 ATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTTTCTGC 383

Db     454 gcacccatggctttgatgggctggaccttgcctggctctaccctgga---c---ggag-- 505
           |||   |  ||||||| || || ||||||| ||||   ||||| |||   |   ||||
Qy     384 GCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGCCAGGGGAGCC 443

Db     506 -------agacaaacagcattttaccaccctaatcaaggaaatgaaggccgaatttataa 558
                  ||||||  |||  || || |||||  |  ||||  ||       |  ||
Qy     444 CTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAATGCCTTCCAGC 503

Db     559 aggaagcccag--c-cagggaaaaagcagctcctgctcagcgcagcactgtctgcgggga 615
           ||||||||||  | |||||||  | |  || || || || |||||  |  | || |||
Qy     504 AGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCCAGCTGGGC 563

Db     616 aggtcaccattgacagcagctatgacattgccaagatatcccaacacctggatttcatta 675
           ||  |   | ||     | || ||  | | ||| ||  |||| ||||||||||  | |
Qy     564 AGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATTTTGTCA 623

Db     676 gcatcatgacctacgattttcatggagcctggcgtgggaccacaggccatcacagtcccc 735
           | | ||| |||||||| || |||||  | |||      |  ||| || ||| |||| ||||
Qy     624 ACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACATAACAGCCCCC 683

Db     736 tgttccgaggtcaggaggatgcaagtcctgacagattcagcaacactgactatgctgtgg 795
           | | |    | || || ||     || | |       |  ||||   ||   |||||||
Qy     684 TCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGATGCTGCTGTGC 743

Db     796 ggtacatgttgaggctgggggctcctgccagtaagctggtgatgggcatccccaccttcg 855
           |   | ||  |  |||| | |||||||| |||||| |  | ||||| || |||| ||
Qy     744 AACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATGCCTACCTACG 803

Db     856 ggaggagcttcactctggcttcttc-tgagactg--gtgttggagccccaatctcaggac 912
           |  |   |||||| |||| || || | ||||    | || || ||||||  | ||||
Qy     804 GACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCAGCCACAGGGT 863

Db     913 cgggaattccaggccggttcaccaaggaggcagggacccttgcctactatgagatctgtg 972
           | || | |||||||| |||||||||| | ||||| || |||||||||||  ||||
Qy     864 CTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTATGAAGTCTGCT 923

Db     973 acttcctccgcggagccacagtccatagaaccctcggccagcaggtcccctatgccacca 1032
           ||      | || |||||    || |||| ||  |  ||| |||| |||||   |  |
Qy     924 CCTGGAAG-G-GG-GCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCCTACATCTTCC 980

Db    1033 agggcaaccagtgggtaggatacgacgaccaggaaagcgtcaaaagcaaggtgcagtacc 1092
           || |||||||||||| || |  || ||   ||| ||| |||||| ||||||    || |
Qy     981 GGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAGGTCAGCTATC 1040

Db    1093 tgaaggataggcagctggcaggcgccatggtatgggccctggacctggatgacttccagg 1152
           ||||| | | |   |||  || |||||||| ||||| |||||| | ||||||||    |
Qy    1041 TGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGATGACTTTGCCG 1100

Db    1153 gctccttctgcggccagg 1170
           ||| || ||||  |||||
Qy    1101 GCTTCTCCTGCAACCAGG 1118


RESULT    3
LOCUS       HSU49835     1418 bp    mRNA            PRI       25-JUL-1996
DEFINITION  Human YKL-39 precursor mRNA, complete cds.
ACCESSION   U49835
NID         g1457940
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1418)
  AUTHORS   Hu,B., Trinh,K., Figueira,W.F. and Price,P.A.
  TITLE     Isolation and sequence of a novel human chondrocyte protein related
            to mammalian members of the chitinase protein family
  JOURNAL   J. Biol. Chem. 271 (32), 19415-19420 (1996)
  MEDLINE   96325055
REFERENCE   2  (bases 1 to 1418)
  AUTHORS   Price,P.A.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1996) Paul A. Price, Biology, 0322, University of
            California at San Diego, 9500 Gilman Drive, La Jolla, CA
            92093-0322, USA
FEATURES             Location/Qualifiers
     source          1..1418
                     /organism="Homo sapiens"
                     /tissue_type="articular cartilage"
     CDS             36..1193
                     /note="similar to human glycoprotein encoded by Genbank
                     Accession Number M80927, human chitotriosidase precursor
                     encoded by Genbank Accession Number U29615, human
                     oviductal glycoprotein encoded by Genbank Accession Number
                     U09550, and to mouse secretory protein YM-1 precursor
                     encoded by Genbank Accession Number S27879"
                     /codon_start=1
                     /product="YKL-39 precursor"
                     /db_xref="PID:g1457941"
                     /translation="MDQKSLWAGVVVLLLLQGGSAYKLVCYFTNWSQDRQEPGKFTPE
                     NIDPFLCSHLIYSFASIENNKVIIKDKSEVMLYQTINSLKTKNPKLKILLSIGGYLFG
                     SKGFHPMVDSSTSRLEFINSIILFLRNHNFDGLDVSWIYPDQKENTHFTVLIHELAEA
                     FQKDFTKSTKERLLLTAGVSAGRQMIDNSYQVEKLAKDLDFINLLSFDFHGSWEKPLI
                     TGHNSPLSKGWQDRGPSSYYNVEYAVGYWIHKGMPSEKVVMGIPTYGHSFTLASAETT
                     VGAPASGPGAAGPITESSGFLAYYEICQFLKGAKITRLQDQQVPYAVKGNQWVGYDDV
                     KSMETKVQFLKNLNLGGAMIWSIDMDDFTGKSCNQGPYPLVQAVKRSLGSL"
     sig_peptide     36..98
     mat_peptide     99..1190
                     /product="YKL-39"
     polyA_site      1418
                     /note="18 A nucleotides"
BASE COUNT      370 a    357 c    337 g    354 t
ORIGIN

  Query Match             9.7%;  Score 159;  DB 74;  Length 1418;
  Best Local Similarity  63.0%;  Pred. No. 5.02e-128;
  Matches   711;  Conservative    0;  Mismatches 390;  Indels  27;  Gaps  10;

Db      53 ctgggcaggtgtagtggtcttgctgcttctccagggaggatctgcctacaaactggtttg 112
           |||||||||| |  ||||| ||||| |  |||   | || |||||    |||||||| ||
Qy      19 CTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTG 78

Db     113 ctactttaccaactggtcccaggaccggcaggaaccaggaaaattcacccctgagaatat 172
           ||||| |||||||| ||||| || | ||||     |     |||  ||  || |  |
Qy      79 CTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTT 138

Db     173 tgacccctttcctatgctctcatctcatctattcattcgccagcatcgaaaacaacaaggt 232
           ||||||  ||| ||| | || |||||||  | |||||  ||||    ||| || || |
Qy     139 GGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCT 198

Db     233 tatcatcaaggacaagagtgaagtgatgctctaccagaccatcaacagtctcaaaaccaa 292
           | || ||  ||   || ||| | ||  |||||||||    ||||  | || || |  |
Qy     199 GAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGAT 258

Db     293 gaatcccaaactgaaaattctcttgtccattggagggtacctgtttggttccaaagggtt 352
           ||||||||| ||||| |  || ||  |||| ||||| |     || ||  |  |   |||
Qy     259 GAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCGGCACTCAGAAGTT 318

Db     353 ccaccctatggtggattcttctacatcacgcttggaattcattaactccataatcctgtt 412
           |     ||||||  |   |  | |     ||   |   ||  | |||||    |||  |||
Qy     319 CACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTT 378

Db     413 tctgaggaaccataactttgatggactggatgtaagctggatctacccag--atc-agaa 469
           |||| | ||  | | |||||| || || ||  |   ||||   |||||||  | | ||
Qy     379 TCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGCCAGGG 438

Db     470 -agaa-------a-acactcat---ttcactgtgctgattcatgagttagcagaagcctt 517
            ||         | |||    |    |||||    ||| | || || || ||  | |||||
Qy     439 GAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAATGCCTT 498

Db     518 tcagaaggacttcacaaaatccaccaaggaaaggcttctcttgactgcgggcgtatctgc 577
           ||| ||||   |   |  ||    |||||| | |||||  ||| ||| |  ||  | ||
Qy     499 CCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCCAGC 558

Db     578 agggaggcaaatgattgataacagctatcaagttgagaaactggcaaaagatctggattt 637
           |||  |       | |||    | ||  | || || ||| | ||  |  | ||||||||
Qy     559 TGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATTT 618

Db     638 catcaacctcctgtcctttgacttccatgggtcttgggaaaagccccttatcactggcca 697
           |||||||  || |||  |||||||||| |||||||| |||        |||| || ||
Qy     619 TGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGG------TCACGGGACA 672

Db     698 caacagccctctgagcaaggggtggcaggacagagggccaagctcctactacaatgtgga 757
           ||||||||  ||   |||| ||    | || || ||  ||    ||  |  ||| |||||
Qy     673 TAACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGA 732

Db     758 atatgctgtggggtactggatacataagggaatgccatcagagaaggtggtcatgggcat 817
           |||||||    | ||| | || ||||| |  ||  |    ||| || || | |||||
Qy     733 TGCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCAT 792

Db     818 ccccacatatgggcactccttcacactggcctc-tg--cagaaaccaccgtgggggcccc 874
           || || || || | ||||||||||||||||||| |   |||| |||| ||||||||||
Qy     793 GCCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCC 852

Db     875 tgcctctggccctggagctgctggacccatcacagagtcttcaggcttcctggcctatta 934
           ||| | ||  |||| || | || ||| |||| ||  |||  | |||||||| ||
Qy     853 AGCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTA 912

Db     935 tgagatctgccagttcctgaaaggagccaagatcacgcgcctccaggatcagcaggttcc 994
           |||  |||||   | |  ||| || |||  |    | |  ||||||||||| |||| ||
Qy     913 TGAAGTCTGC---TCCTGGAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCC 969

Db     995 ctacgcagtcaaggggaaccagtgggtgggctatgatgatgtgaagagtatggagaccaa 1054
           ||||    ||  ||  |||||||||||||||| |||||||||| |||| |  | |||||
Qy     970 CTACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAA 1029

Db    1055 ggttcagttcttaaagaatttaaacctgggaggagccatgatctggtctattgacatgga 1114
           |||    |   | ||| |        ||||| || |||||| ||||| |  | ||| | ||
Qy    1030 GGTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGA 1089

Db    1115 tgacttcactggcaaatcctgcaaccagggcccttaccctcttgtcca 1162
           ||||||  | |||  |||||||||||||||  ||||| ||  ||||
Qy    1090 TGACTTTGCCGGCTTCTCCTGCAACCAGGGCCGATACCCCCTCATCCA 1137


RESULT    4
LOCUS       HSU58514     1434 bp    mRNA            PRI       24-JUL-1996
DEFINITION  Human chitinase precursor (HUMTCHIT) mRNA, exon 1a form, complete
            cds.
ACCESSION   U58514
NID         g1439565
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1434)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1434)
  AUTHORS   Grossman,A., Matsuyama,T., Baker,E., Waterhouse,P., Sutherland,G.R.
            and Mak,T.W.
  TITLE     Cloning of a novel lymphoid restricted human chitinase and
            localization to 1p13.3
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1434)
  AUTHORS   Grossman,A., Matsuyama,T., Baker,E., Waterhouse,P., Sutherland,G.R.
            and Mak,T.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1996) Medical Biophysics, Ontario Cancer
            Institute, Princess Margaret Hospital, 610 University Avenue,
            Toronto, Ontario M5G 2M9, Canada
```

```
FEATURES             Location/Qualifiers
     source          1..1434
                     /organism="Homo sapiens"
                     /map="1p13.3"
                     /chromosome="1"
                     /tissue_type="Soares placenta"
                     /dev_stage="8 to 9 weeks"
                     /clone="EST 257753"
     5'UTR           1..36
                     /gene="HUMTCHIT"
     exon            1..76
                     /gene="HUMTCHIT"
                     /note="exon 1a; first alternatively splice exon 1;
                     contains a Kozak consensus initiator methionine"
                     /number=1
     sig_peptide     37..114
                     /gene="HUMTCHIT"
     CDS             37..1209
                     /gene="HUMTCHIT"
                     /note="first alternatively spliced form, see GenBank
                     Accession Number U58515 for the second alternatively
                     spliced form"
                     /codon_start=1
                     /product="chitinase precursor"
                     /db_xref="PID:g1439566"
                     /translation="MGATTMDQKSLWAGVVVLLLLQGGSAYKLVCYFTNWSQDRQEPG
                     KFTPENIDPFLCSHLIYSFASIENNKVIIKDKSEVMLYQTINSLKTKNPKLKILLSIG
                     GYLFGSKGFHPMVDSSTSRLEFINSIILFLRNHNFDGLDVSWIYPDQKENTHFTVLIH
                     ELAEAFQKDFTKSTKERLLLTAGVSAGRQMIDNSYQVEKLAKDLDFINLLSFDFHGSW
                     EKPLITGHNSPLSKGWQDRGPSSYYNVEYAVGYWIHKGMPSEKVVMGIPTYGHSFTLA
                     SAETTVGAPASGPGAAGPITESSGFLAYYEICQFLKGAKITRLQDQQVPYAVKGNQWV
                     GYDDVKSMETKVQFLKNLNLGGAMIWSIDMDDFTGKSCNQGPYPLVQAVKRSLGSL"
     exon            77..106
                     /gene="HUMTCHIT"
                     /number=2
     exon            107..309
                     /gene="HUMTCHIT"
                     /number=3
     mat_peptide     115..1206
                     /gene="HUMTCHIT"
                     /note="first alternatively spliced form, see GenBank
                     Accession Number U58515 for the second alternatively
                     spliced form"
                     /product="chitinase"
     exon            310..365
                     /gene="HUMTCHIT"
                     /number=4
     exon            366..516
                     /gene="HUMTCHIT"
                     /number=5
     exon            517..643
                     /gene="HUMTCHIT"
                     /number=6
     exon            644..771
                     /gene="HUMTCHIT"
                     /number=7
     exon            772..954
                     /gene="HUMTCHIT"
                     /number=8
     exon            955..1071
                     /gene="HUMTCHIT"
                     /number=9
     exon            1072..1433
                     /gene="HUMTCHIT"
                     /number=10
     3'UTR           1210..1433
                     /gene="HUMTCHIT"
     polyA_signal    1407..1412
                     /gene="HUMTCHIT"
     polyA_site      1434
                     /gene="HUMTCHIT"
                     /note="32 A nucleotides"
BASE COUNT      375 a    360 c    342 g    357 t
```

```
ORIGIN

  Query Match             9.7%;  Score 159;  DB 74;  Length 1434;
  Best Local Similarity  63.0%;  Pred. No. 5.02e-128;
  Matches   711;  Conservative    0;  Mismatches 390;  Indels  27;  Gaps  10;

Db      69 ctgggcaggtgtagtggtcttgctgcttctccagggaggatctgcctacaaactggtttg 128
           |||||||||| | ||||| ||||| | |||  | || |||||   ||||||||  ||
Qy      19 CTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTG 78

Db     129 ctactttaccaactggtcccaggaccggcaggaaccaggaaaattcacccctgagaatat 188
           |||||| |||||||||| ||||| || | |||||     |     |||   ||  || | |
Qy      79 CTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTT 138

Db     189 tgacccctcctatgctctcatctcatctattcattcgccagcatcgaaaacaacaaggt 248
           ||||||  ||| ||| | || ||||||||  | |||||  ||||    ||| || || |
Qy     139 GGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCT 198

Db     249 tatcatcaaggacaagagtgaagtgatgctctaccagaccatcaacagtctcaaaaccaa 308
           | || ||  ||   || ||| | ||  |||||||||    ||||  | || || |  |
Qy     199 GAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGAT 258

Db     309 gaatcccaaactgaaaattctcttgtccattggagggtacctgtttggttccaaagggtt 368
           ||||||||| ||||| |  || ||  |||| ||||| |     || ||  |  |   |||
Qy     259 GAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCGGCACTCAGAAGTT 318

Db     369 ccaccctatggtggattcttctacatcacgcttggaattcattaactccataatcctgtt 428
           |     |||||| |   |  | |     ||   |   ||  | |||||    |||  |||
Qy     319 CACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTT 378

Db     429 tctgaggaaccataactttgatggactggatgtaagctggatctacccag--atc-agaa 485
           |||| | ||  | | |||||| || || ||  |   ||||   |||||||  | | ||
Qy     379 TCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGCCAGGG 438

Db     486 -agaa-------a-acactcat---ttcactgtgctgattcatgagttagcagaagcctt 533
           ||          | |||   |    |||||    ||| | || || || ||  | |||||
Qy     439 GAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAATGCCTT 498

Db     534 tcagaaggacttcacaaaatccaccaaggaaaggcttctcttgactgcgggcgtatctgc 593
           ||| ||||   |   |  ||    |||||| | |||||  ||| ||| |  ||  | ||
Qy     499 CCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCCAGC 558

Db     594 agggaggcaaatgattgataacagctatcaagttgagaaactggcaaaagatctggattt 653
           |||  |      | |||    | ||  | || || ||| | ||  |  | |||||||||
Qy     559 TGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATTT 618

Db     654 catcaacctcctgtcctttgacttccatgggtcttgggaaaagccccttatcactggcca 713
           |||||||  || |||  |||||||||| |||||||| |||      |||| || ||
Qy     619 TGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGG------TCACGGGACA 672

Db     714 caacagccctctgagcaaggggtggcaggacagagggccaagctcctactacaatgtgga 773
           |||||||| ||   |||| ||    | || || ||  ||    ||  |  ||| |||||
Qy     673 TAACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGA 732

Db     774 atatgctgtggggtactggatacataagggaatgccatcagagaaggtggtcatgggcat 833
           |||||||   | ||| | || ||||| |  ||  |    ||| || || | |||||
Qy     733 TGCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCAT 792

Db     834 ccccacatatgggcactccttcacactggcctc-tg--cagaaaccaccgtgggggcccc 890
           || || || || | ||||||||||||||||||  |   |||| ||||  ||||||||||
Qy     793 GCCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCC 852

Db     891 tgcctctggccctggagctgctggacccatcacagagtcttcaggcttcctggcctatta 950
           ||| | ||  |||| || | || ||| |||| ||      |||  | |||||||| ||
Qy     853 AGCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTA 912

Db     951 tgagatctgccagttcctgaaaggagccaagatcacgcggctccaggatcagcaggttcc 1010
           |||  |||||   | |  ||| || ||||  |    | |  ||||||||||| |||| ||
Qy     913 TGAAGTCTGC---TCCTGGAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCC 969

Db    1011 ctacgcagtcaaggggaaccagtgggtgggctatgatgatgtgaagagtatggagaccaa 1070
           ||||    ||  ||  |||||||||||||||| ||||||||| |||| |  | |||||
Qy     970 CTACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAA 1029
```

```
Db    1071 ggttcagttcttaaagaatttaaacctgggaggagccatgatctggtctattgacatgga 1130
           |||    |   | ||| |       ||||| || |||||| ||||| |  | ||| | ||
Qy    1030 GGTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGA 1089

Db    1131 tgacttcactggcaaatcctgcaaccagggcccttaccctcttgtcca 1178
           ||||||  | |||   ||||||||||||||||  ||||| ||  ||||
Qy    1090 TGACTTTGCCGGCTTCTCCTGCAACCAGGGCCGATACCCCCTCATCCA 1137


RESULT    5
LOCUS       HSU58515     1500 bp    mRNA            PRI       24-JUL-1996
DEFINITION  Human chitinase (HUMTCHIT) mRNA, exon 1b form, partial cds.
ACCESSION   U58515
NID         g1439567
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1500)
  AUTHORS   Grossman,A., Matsuyama,T., Baker,E., Waterhouse,P., Sutherland,G.R.
            and Mak,T.W.
  TITLE     Cloning of a novel lymphoid restricted human chitinase and
            localization to 1p13.3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1500)
  AUTHORS   Grossman,A., Matsuyama,T., Baker,E., Waterhouse,P., Sutherland,G.R.
            and Mak,T.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1996) Medical Biophysics, Ontario Cancer
            Institute, Princess Margaret Hospital, 610 University Avenue,
            Toronto, Ontario M5G 2M9, Canada
FEATURES             Location/Qualifiers
     source          1..1500
                     /organism="Homo sapiens"
                     /map="1p13.3"
                     /chromosome="1"
                     /clone_lib="Stratgene Jurkat T-cell cDNA library in Lambda
                     Zap II; Clontech thymus polyA+ RNA; Stratagene placental
                     genomic library in Lambda FIX II; activated T-cell plasmid
                     library (Dr. Brian Seed)"
     exon            1..142
                     /gene="HUMTCHIT"
                     /note="exon 1b; second alternatively spliced exon 1"
                     /number=1
     CDS             <1..1275
                     /evidence=not_experimental
                     /gene="HUMTCHIT"
                     /note="submitter can find no appropriate Kozak initiator
                     methionine, despite the open reading frame continuing
                     upstream, and is certain that there are no other exons 5'
                     as determined by RACE and genomic sequencing; use of a Met
                     further downstream as the initiator Met would result in
                     the deletion of amino acids conserved in all other family
                     members; submitter hypothesizes that an Ile is the
                     initiator codon for this form; second alternatively
                     spliced form; see GenBank Accession Number U58514 for the
                     first alternatively spliced form"
                     /codon_start=1
                     /product="chitinase"
                     /db_xref="PID:g1439568"
                     /translation="SQQLWLGSPDEVWLYLVCEHTHIFTAIIWDSRTRFGSTDFSFHP
                     SIAGVVVLLLLQGGSAYKLVCYFTNWSQDRQEPGKFTPENIDPFLCSHLIYSFASIEN
                     NKVIIKDKSEVMLYQTINSLKTKNPKLKILLLSIGGYLFGSKGFHPMVDSSTSRLEFIN
                     SIILFLRNHNFDGLDVSWIYPDQKENTHFTVLIHELAEAFQKDFTKSTKERLLLTAGV
                     SAGRQMIDNSYQVEKLAKDLDFINLLSFDFHGSWEKPLITGHNSPLSKGWQDRGPSSY
                     YNVEYAVGYWIHKGMPSEKVVMGIPTYGHSFTLASAETTVGAPASGPGAAGPITESSG
                     FLAYYEICQFLKGAKITRLQDQQVPYAVKGNQWVGYDDVKSMETKVQFLKNLNLGGAM
                     IWSIDMDDFTGKSCNQGPYPLVQAVKRSLGSL"
     exon            143..172
                     /gene="HUMTCHIT"
                     /number=2
     exon            173..375
                     /gene="HUMTCHIT"
                     /number=3
     exon            376..431
                     /gene="HUMTCHIT"
                     /number=4
     exon            432..582
                     /gene="HUMTCHIT"
                     /number=5
     exon            583..709
                     /gene="HUMTCHIT"
                     /number=6
     exon            710..847
                     /gene="HUMTCHIT"
                     /number=7
     exon            848..1020
                     /gene="HUMTCHIT"
                     /number=8
     exon            1021..1137
                     /gene="HUMTCHIT"
                     /number=9
     exon            1138..1277
                     /gene="HUMTCHIT"
                     /number=10
     3'UTR           1276..1499
                     /gene="HUMTCHIT"
     exon            1278..1499
                     /gene="HUMTCHIT"
                     /number=11
     polyA_signal    1473..1478
                     /gene="HUMTCHIT"
     polyA_site      1500
                     /gene="HUMTCHIT"
                     /note="18 A nucleotides"
BASE COUNT      385 a    380 c    351 g    384 t
ORIGIN

  Query Match           9.6%;  Score 157;  DB 74;  Length 1500;
  Best Local Similarity  63.0%;  Pred. No. 6.03e-126;
  Matches   708;  Conservative    0;  Mismatches 389;  Indels  27;  Gaps  10;

Db     139 gcaggtgtagtggtcttgctgcttctccagggaggatctgcctacaaactggtttgctac 198
           |||||| |  ||||| ||||| |  |||   | || |||||    |||||||| ||||||
Qy      23 GCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTGCTAC 82

Db     199 tttaccaactggtcccaggaccggcaggaaccaggaaaattcacccctgagaatattgac 258
           || |||||||||| ||||| || | |||||      |     |||   ||  || |  | |||
Qy      83 TTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGAC 142

Db     259 cccttcctatgctctcatctcatctattcattcgccagcatcgaaaacaacaaggttatc 318
           |||  ||| ||| | || |||||||  | |||||  ||||    ||| || || | | |
Qy     143 CCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCTGAGC 202

Db     319 atcaaggacaagagtgaagtgatgctctaccagaccatcaacagtctcaaaaccaagaat 378
           | ||  ||   || ||| | ||  |||||||||    ||||  | || || |  | ||||
Qy     203 ACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGATGAAT 262

Db     379 cccaaactgaaaattctcttgtccattggagggtacctgtttggttccaaagggttccac 438
           ||||| ||||| |  || ||  |||| |||||  |     || ||  |  |   ||||
Qy     263 CCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCGGCACTCAGAAGTTCACA 322

Db     439 cctatggtggattcttctacatcacgcttggaattcattaactccataatcctgtttctg 498
             |||||| |   |  | |     ||   |   ||  | |||||    |||  |||||||
Qy     323 GATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTTTCTG 382

Db     499 aggaaccataactttgatggactggatgtaagctggatctacccag--atc-agaa-aga 554
             | ||  | | |||||| || || ||  |   ||||   |||||||  | | ||   ||
Qy     383 CGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGCCAGGGGAGC 442

Db     555 a-------a-acactcat---ttcactgtgctgattcatgagttagcagaagcctttcag 603
                   | |||   |    |||||    ||| | || || || ||  | ||||| |||
Qy     443 CCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAATGCCTTCCAG 502
```

```
Db    604 aaggacttcacaaaatccaccaaggaaaggcttctcttgactgcgggcgtatctgcaggg 663
Qy    503 CAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCCAGCTGGG 562

Db    664 aggcaaatgattgataacagctatcaagttgagaaactggcaaaagatctggatttcatc 723
Qy    563 CAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATTTTGTC 622

Db    724 aacctcctgtcctttgacttccatgggtcttgggaaaagccccttatcactggccacaac 783
Qy    623 AACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGG------TCACGGGACATAAC 676

Db    784 agccctctgagcaaggggtggcaggacagagggccaagctcctactacaatgtggaatat 843
Qy    677 AGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGATGCT 736

Db    844 gctgtggggtactggatacataagggaatgccatcagagaaggtggtcatgggcatcccc 903
Qy    737 GCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATGCCT 796

Db    904 acatatggacactccttcacactggcctc-tg--cagaaaccaccgtgggggcccctgcc 960
Qy    797 ACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCAGCC 856

Db    961 tctggccctggagctgctggacccatcacagagtcttcaggcttcctggcctattatgag 1020
Qy    857 ACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTATGAA 916

Db   1021 atctgccagttcctgaaaggagccaagatcacgcggctccaggatcagcaggttccctac 1080
Qy    917 GTCTGC---TCCTGGAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCCTAC 973

Db   1081 gcagtcaaggggaaccagtgggtgggctatgatgatgtgaagagtatggagaccaaggtt 1140
Qy    974 ATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAGGTC 1033

Db   1141 cagttcttaaagaatttaaacctgggaggagccatgatctggtctattgacatggatgac 1200
Qy   1034 AGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGATGAC 1093

Db   1201 ttcactggcaaatcctgcaaccagggcccttaccctcttgtcca 1244
Qy   1094 TTTGCCGGCTTCTCCTGCAACCAGGGCCGATACCCCCTCATCCA 1137
```

```
RESULT    6
LOCUS       SSGP38KD     1733 bp    RNA             MAM       28-AUG-1996
DEFINITION  S.scrofa 38kDa heparin-binding glycoprotein.
ACCESSION   Z47803
NID         g634097
KEYWORDS    gp38k; heparin-binding glycoprotein.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Artiodactyla; Suiformes; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 1733)
  AUTHORS   Millis,A.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-1995) Millis A. J., University at Albany,
            Biological Sciences, 1400 Washington Avenue, Albany, NY, USA, 12222
REFERENCE   2  (bases 1 to 1733)
  AUTHORS   Shackelton,L.M., Mann,D.M. and Millis,A.J.
  TITLE     Identification of a 38-kDa heparin-binding glycoprotein (gp38k) in
            differentiating vascular smooth muscle cells as a member of a group
            of proteins associated with tissue remodeling
  JOURNAL   J. Biol. Chem. 270 (22), 13076-13083 (1995)
  MEDLINE   95286589
FEATURES             Location/Qualifiers
     source          1..1733
                     /organism="Sus scrofa"
                     /clone="pBS38k"
                     /dev_stage="adult"
                     /tissue_type="smooth muscle"
                     /clone_lib="lambda ZAP II"
     CDS             67..1218
                     /codon_start=1
                     /product="38 kDa heparin-binding glycoprotein"
                     /db_xref="PID:g634098"
                     /translation="MGLRVAQTGFVALVLLQSCAAYKLVCYYTSWSQYREGDGSCFPD
                     AIDPFLCTHIIYSFANISNNEIDTLEWNDVTLYDTLNTLKNRNPNLKTLLSVGGWNFG
                     SQRFSKIASNTQSRRTFIKSVPPFLRTHGFDGLDLAWISPGRRDKRHLTTLVKEMKAE
                     FVREALPGTERLLLSGAVSAGKVAIDRGYDIAQISQHLDFISLLTYDFHGAWRQTTGH
                     HSPLFRGQGDASSDRFSNADYAVSYVLRLGAPANKLVMGIPTFGRSFTLASSKTDVGA
                     PASGPGIPGRFTKEKGILAYYEICDFLQGATVRRPLGQQVPYATKGNQWVGYDDQESV
                     KNKAKYLKSRQLAGAMVWTLDLDDFRGNFCGQNLRFPLTSAIKDVLAAA"
BASE COUNT      408 a    507 c    463 g    355 t
ORIGIN
```

```
  Query Match           9.5%;  Score 156;  DB 48;  Length 1733;
  Best Local Similarity  62.5%;  Pred. No. 6.59e-125;
  Matches   688;  Conservative    0;  Mismatches 388;  Indels  24;  Gaps   9;

Db     79 gtggctcagacaggttttgtggccctggtgctgctccagagctgtgctgcctacaagctg 138
Qy     14 GTGGCCTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTG 73

Db    139 gtttgctactataccagctggtctcagtaccgggagggtgatgggagctgcttcccagat 198
Qy     74 GTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAG 133

Db    199 gccatcgacccctcctctgcacccacatcatctacagctttgccaacataagcaacaat 258
Qy    134 GACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCAC 193

Db    259 gagattgacaccttggagtggaatgatgtgacgctctatgacacactgaacacactcaag 318
Qy    194 CAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAG 253

Db    319 aacaggaaccccaacctgaagaccctcctgtctgttggaggatggaactttggttctcaa 378
Qy    254 AAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCGGCACTCAG 313

Db    379 agattttccaaaatagcctccaacactcagagtcgcaggactttcatcaagtcggtgcca 438
Qy    314 AAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATC 373

Db    439 ccattcctgcggacccatggctttgatggactggacctagcctggatctcccctgg--gc 496
Qy    374 AGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGC 433

Db    497 ----ggag---------agacaagcggcatctcaccactctagtcaaggagatgaaggct 543
Qy    434 CAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAAT 493

Db    544 gagtttgtaagggaagccctg-cct--ggaacagagcggctcctgctcagcggagcggtg 600
Qy    494 GCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTT 553

Db    601 tctgcagggaaggtcgccattgacagaggctatgacatcgcccagatatctcaacacctg 660
Qy    554 CCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTG 613

Db    661 gacttcatcagccttttgacctatgacttccacggtgcctggcgccaaactacaggacac 720
Qy    614 GATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACAT 673

Db    721 cacagtcccctgttccgaggccaggggatgcaagttccgacagattcagcaatgcagac 780
Qy    674 AACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGAT 733

Db    781 tatgctgtgagctatgtgctgaggctgggggccccagccaataagctggtgatgggtatc 840
Qy    734 GCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATG 793

Db    841 cccacttttgggaggagcttcactttggcctcttcca-agacag-ac-gtgggagccccg 897
```

```
Qy    794 CCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCA 853

Db    898 gcatcagggccaggaataccaggccggttcaccaaggagaaagggatccttgcctactat 957
Qy    854 GCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTAT 913

Db    958 gagatctgtgacttcctccaaggagccacagtccgtagacccccttggccagcaggtcccc 1017
Qy    914 GAAGTCTG---CTCCTGGAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCC 970

Db   1018 tatgccaccaagggcaaccagtgggtggggtatgatgaccaggagagcgtcaaaaacaag 1077
Qy    971 TACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAG 1030

Db   1078 gcaaagtacctgaagagcaggcagctggctggtgctatggtgtggaccctggacttggat 1137
Qy   1031 GTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGAT 1090

Db   1138 gactttcggggcaacttctg 1157
Qy   1091 GACTTTGCCGGCTTCTCCTG 1110


RESULT    7
LOCUS       SSU19900     1733 bp    mRNA            MAM       02-FEB-1996
DEFINITION  Sus scrofa 38kDa heparin-binding glycoprotein mRNA, complete cds.
ACCESSION   U19900
NID         g643470
KEYWORDS    .
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Artiodactyla; Suiformes; Suina; Suidae; Sus.
REFERENCE   1  (sites)
  AUTHORS   Shackelton,L.M., Mann,D.M. and Millis,A.J.
  TITLE     Identification of a 38-kDa heparin-binding glycoprotein (gp38k) in
            differentiating vascular smooth muscle cells as a member of a group
            of proteins associated with tissue remodeling
  JOURNAL   J. Biol. Chem. 270 (22), 13076-13083 (1995)
  MEDLINE   95286589
REFERENCE   2  (bases 1 to 1733)
  AUTHORS   Millis,A.J.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JAN-1995) Albert J.T. Millis, Biological Sciences,
            University at Albany, 1400 Washington Avenue, Albany, NY 12222, USA
FEATURES             Location/Qualifiers
     source          1..1733
                     /organism="Sus scrofa"
                     /note="in Lambda Zap II"
                     /clone="pBS38k"
                     /tissue_type="smooth muscle"
                     /dev_stage="adult"
     CDS             67..1218
                     /note="gp38k"
                     /codon_start=1
                     /product="38kDa heparin-binding glycoprotein"
                     /db_xref="PID:g643471"
                     /translation="MGLRVAQTGFVALVLLQSCAAYKLVCYYTSWSQYREGDGSCFPD
                     AIDPFLCTHIIYSFANISNNEIDTLEWNDVTLYDTLNTLKNRNPNLKTLLSVGGWNFG
                     SQRFSKIASNTQSRRTFIKSVPPFLRTHGFDGLDLAWISPGRRDKRHLTTLVKEMKAE
                     FVREALPGTERLLLSGAVSAGKVAIDRGYDIAQISQHLDFISLLTYDFHGAWRQTTGH
                     HSPLFRGQGDASSDRFSNADYAVSYVLRLGAPANKLVMGIPTFGRSFTLASSKTDVGA
                     PASGPGIPGRFTKEKGILAYYEICDFLQGATVRRPLGQQVPYATKGNQWVGYDDQESV
                     KNKAKYLKSRQLAGAMVWTLDLDDFRGNFCGQNLRFPLTSAIKDVLAAA"
BASE COUNT      408 a    507 c    463 g    355 t
ORIGIN

  Query Match           9.5%;  Score 156;  DB 48;  Length 1733;
  Best Local Similarity  62.5%;  Pred. No. 6.59e-125;
  Matches   688;  Conservative    0;  Mismatches 388;  Indels  24;  Gaps   9;

Db     79 gtggctcagacaggttttgtggccctggtgctgctccagagctgtgctgcctacaagctg 138
Qy     14 GTGGCCTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTG 73

Db    139 gtttgctactataccagctggtctcagtaccgggagggtgatgggagctgcttcccagat 198
Qy     74 GTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAG 133

Db    199 gccatcgacccccttcctctgcacccacatcatctacagctttgccaacataagcaacaat 258
Qy    134 GACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCAC 193

Db    259 gagattgacaccttggagtggaatgatgtgacgctctatgacacactgaacacactcaag 318
Qy    194 CAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAG 253

Db    319 aacaggaaccccaacctgaagaccctcctgtctgttggaggatggaactttggttctcaa 378
Qy    254 AAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCGGCACTCAG 313

Db    379 agattttccaaaatagcctccaacactcagagtcgcaggactttcatcaagtcggtgcca 438
Qy    314 AAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATC 373

Db    439 ccattcctgcggacccatggctttgatggactggacctagcctggatctcccctgg--gc 496
Qy    374 AGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGC 433

Db    497 ----ggag---------agacaagcggcatctcaccactctagtcaaggagatgaaggct 543
Qy    434 CAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAAT 493

Db    544 gagtttgtaagggaagccctg-cct--ggaacagagcggctcctgctcagcggagcggtg 600
Qy    494 GCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTT 553

Db    601 tctgcagggaaggtcgccattgacagaggctatgacatcgcccagatatctcaacacctg 660
Qy    554 CCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTG 613

Db    661 gacttcatcagccttttgacctatgacttccacggtgcctggcgccaaactacaggacac 720
Qy    614 GATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACAT 673

Db    721 cacagtcccctgttccgaggccaggggatgcaagttccgacagattcagcaatgcagac 780
Qy    674 AACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGAT 733

Db    781 tatgctgtgagctatgtgctgaggctgggggccccagccaataagctggtgatgggtatc 840
Qy    734 GCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATG 793

Db    841 cccacttttgggaggagcttcactttggcctcttcca-agacag-ac-gtgggagccccg 897
Qy    794 CCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCA 853

Db    898 gcatcagggccaggaataccaggccggttcaccaaggagaaagggatccttgcctactat 957
Qy    854 GCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTAT 913

Db    958 gagatctgtgacttcctccaaggagccacagtccgtagacccccttggccagcaggtcccc 1017
Qy    914 GAAGTCTG---CTCCTGGAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCC 970

Db   1018 tatgccaccaagggcaaccagtgggtggggtatgatgaccaggagagcgtcaaaaacaag 1077
Qy    971 TACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAG 1030

Db   1078 gcaaagtacctgaagagcaggcagctggctggtgctatggtgtggaccctggacttggat 1137
Qy   1031 GTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGAT 1090

Db   1138 gactttcggggcaacttctg 1157
Qy   1091 GACTTTGCCGGCTTCTCCTG 1110
```

```
RESULT    8
ID   SSGP38KD   standard; RNA; MAM; 1733 BP.
AC   Z47803;
NI   g634097
DT   19-JAN-1995 (Rel. 42, Created)
DT   28-AUG-1996 (Rel. 49, Last updated, Version 6)
DE   S.scrofa 38kDa heparin-binding glycoprotein.
KW   gp38k; heparin-binding glycoprotein.
OS   Sus scrofa (domestic pig)
OC   Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC   Theria; Eutheria; Artiodactyla; Suiformes; Suidae.
RN   [2]
RP   1-1733
RA   Millis A.J.;
RT   ;
RL   Submitted (19-JAN-1995) to the EMBL/GenBank/DDBJ databases.
RL   Millis A. J., University at Albany, Biological Sciences, 1400
RL   Washington Avenue, Albany, NY, USA, 12222
RN   [3]
RP   1-1733
RX   MEDLINE; 95286589.
RA   Shackelton L.M., Mann D.M., Millis A.J.;
RT   "Identification of a 38-kDa heparin-binding glycoprotein (gp38k)
RT   in differentiating vascular smooth muscle cells as a member of a
RT   group of proteins associated with tissue remodeling";
RL   J. Biol. Chem. 270:13076-13083(1995).
FH   Key             Location/Qualifiers
FH
FT   source          1..1733
FT                   /organism="Sus scrofa"
FT                   /clone="pBS38k"
FT                   /dev_stage="adult"
FT                   /tissue_type="smooth muscle"
FT                   /clone_lib="lambda ZAP II"
FT   CDS             67..1218
FT                   /product="38 kDa heparin-binding glycoprotein"
FT                   /db_xref="PID:g634098"
FT                   /translation="MGLRVAQTGFVALVLLQSCAAYKLVCYYTSWSQYREGDGSCFPD
  A
FT                   IDPFLCTHIIYSFANISNNEIDTLEWNDVTLYDTLNTLKNRNPNLKTLLSVGGWNFGS
  Q
FT                   RFSKIASNTQSRRTFIKSVPPFLRTHGFDGLDLAWISPGRRDKRHLTTLVKEMKAEFV
  R
FT                   EALPGTERLLLSGAVSAGKVAIDRGYDIAQISQHLDFISLLTYDFHGAWRQTTGHHSP
  L
FT                   FRGQGDASSDRFSNADYAVSYVLRLGAPANKLVMGIPTFGRSFTLASSKTDVGAPASG
  P
FT                   GIPGRFTKEKGILAYYEICDFLQGATVRRPLGQQVPYATKGNQWVGYDDQESVKNKAK
  Y
FT                   LKSRQLAGAMVWTLDLDDFRGNFCGQNLRFPLTSAIKDVLAAA"
SQ   Sequence 1733 BP; 408 A; 507 C; 463 G; 355 T; 0 other;

  Query Match             9.5%;  Score 156;  DB 15;  Length 1733;
  Best Local Similarity  62.5%;  Pred. No. 6.59e-125;
  Matches   688;  Conservative    0;  Mismatches 388;  Indels  24;  Gaps   9;

Db     79 gtggctcagacaggttttgtggccctggtgctgctccagagctgtgctgcctacaagctg 138
          |||||   | |||||||  ||| |||| || || |||   |  |  ||||    || |||
Qy     14 GTGGCCTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTG 73

Db    139 gtttgctactataccagctggtctcagtaccgggagggtgatgggagctgcttcccagat 198
          || |||||||  |||| |||| | |||||| |  |||| || |   ||| | | ||  |
Qy     74 GTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAG 133

Db    199 gccatcgacccccttcctctgcacccacatcatctacagctttgccaacataagcaacaat 258
          | | | ||||||  ||| |||||||||| ||||||||  ||| ||   ||| | |||| |
Qy    134 GACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCAC 193

Db    259 gagattgacaccttggagtggaatgatgtgacgctctatgacacactgaacacactcaag 318
          || |   ||||   |||||||||||| | ||| |||||  |      | ||    || |||
Qy    194 CAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAG 253

Db    319 aacaggaaccccaacctgaagaccctcctgtctgttggaggatggaactttggttctcaa 378
          || | ||| ||||| ||||||||||||  |  |  | ||||| ||||| || ||  ||||
Qy    254 AAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCGGCACTCAG 313

Db    379 agattttccaaaatagcctccaacactcagagtcgcaggactttcatcaagtcggtgcca 438
          |  ||  |  | || |   |||   |  | | ||   ||| ||  |||| ||||
Qy    314 AAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATC 373

Db    439 ccattcctgcggacccatggctttgatggactggacctagcctggatctcccctgg--gc 496
             || ||||| |   |  |||||||  || || |||||  | ||||   | |||  ||  ||
Qy    374 AGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGC 433

Db    497 ----ggag---------agacaagcggcatctcaccactctagtcaaggagatgaaggct 543
              ||||         |||||||  ||   |||| || || ||  ||||  ||     |
Qy    434 CAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAAT 493

Db    544 gagtttgtaagggaagccctg-cct--ggaacagagcggctcctgctcagcggagcggtg 600
          |  ||      |||||||| | |||  || |  || || || || || || || | ||||||
Qy    494 GCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTT 553

Db    601 tctgcagggaaggtcgccattgacagaggctatgacatcgcccagatatctcaacacctg 660
          | || ||| ||  |    | ||    || || ||  | | | | ||  | ||  |||||
Qy    554 CCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTG 613

Db    661 gacttcatcagccttttgacctatgacttccacggtgcctggcgccaaactacaggacac 720
          || ||  ||| |||| || |||| ||||||| ||  | |||    |    || |||||
Qy    614 GATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACAT 673

Db    721 cacagtcccctgttccgaggccaggggggatgcaagttccgacagattcagcaatgcagac 780
          |||| ||||| | |   | || | ||     || | |       | ||| |  ||
Qy    674 AACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGAT 733

Db    781 tatgctgtgagctatgtgctgaggctgggggccccagccaataagctggtgatgggtatc 840
          |||||||    |   ||||  |  |||| |||| ||||  |||||| |  | || ||
Qy    734 GCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATG 793

Db    841 cccacttttgggaggagcttcactttggcctcttcca-agacag-ac-gtgggagccccg 897
          || || |  ||  |  ||||||  |||||| ||   |||||  |  ||||| |||||
Qy    794 CCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCA 853

Db    898 gcatcagggccaggaataccaggccggttcaccaaggagaaagggatccttgcctactat 957
          ||  ||||| | || |  |||||||  ||||||||||   |||||| || |||||||||
Qy    854 GCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTAT 913

Db    958 gagatctgtgacttcctccaaggagccacagtccgtagacccccttggccagcaggtcccc 1017
          ||  ||||   || |    | || |||||    |  |||  ||  |  ||| |||| |||
Qy    914 GAAGTCTG---CTCCTGGAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCC 970

Db   1018 tatgccaccaagggcaaccagtgggtggggtatgatgaccaggagagcgtcaaaaacaag 1077
          ||   |  | || |||||||||||||||| | ||||||   |||||| |||||| ||||
Qy    971 TACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAG 1030

Db   1078 gcaaagtacctgaagagcaggcagctggctggtgctatggtgtggaccctggacttggat 1137
          |  |  || ||||||   | |   ||||  || || ||||| ||| | |||||||| |||
Qy   1031 GTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGAT 1090

Db   1138 gactttcggggcaacttctg 1157
          ||||||   |||  || |||
Qy   1091 GACTTTGCCGGCTTCTCCTG 1110


RESULT    9
LOCUS       MMBRP39      1616 bp    RNA             ROD       28-NOV-1995
DEFINITION  M.musculus mRNA for BRP39 protein.
ACCESSION   X93035
NID         g1085065
KEYWORDS    brp39 gene; BRP39 protein.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 1616)
  AUTHORS   Morrison,B.W. and Leder,P.
  TITLE     neu and ras initiate murine mammary tumors that share genetic
            markers generally absent in c-myc and int-2-initiated tumors
  JOURNAL   Oncogene 9 (12), 3417-3426 (1994)
  MEDLINE   95060797
REFERENCE   2  (bases 1 to 1616)
  AUTHORS   Morrison,B.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-1995) B.W. Morrison, Dana-Farber Cancer
            Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1..1616
                     /organism="Mus musculus"
                     /strain="FUB/N"
                     /dev_stage="adult"
                     /tissue_type="breast"
                     /cell_line="SMF murine breast tumor"
                     /clone_lib="lambda ZAP II"
     misc_feature    66..128
                     /gene="brp39"
                     /note="leader sequence"
     CDS             66..1211
                     /gene="brp39"
                     /note="homology to glycosyl hydrolases"
                     /codon_start=1
                     /product="BRP39 protein"
                     /db_xref="PID:e211795"
                     /db_xref="PID:g1085066"
                     /translation="MGMRAALTGFAVLMLLQSCSAYKLVCYFTSWSQYREGVGSFLPD
                     AIQPFLCTHIIYSFANISSDNMLSTWEWNDESNYDKLNKLKTRNTNLKTLLSVGGWKF
                     GEKRFSEIASNTERRTAFVRSVAPFLRSYGFDGLDLAWLYPRLRDKQYFSTLIKELNA
                     EFTKEVQPGREKLLLSAALSAGKVAIDTGYDIAQIAQHLDFINLMTYDFHGVWRQITG
                     HHSPLFQGQKDTRFDRYSNVNYAVQYMIRLGAQASKLLMGIPTFGKSFTLASSENQLG
                     APISGEGLPGRFTKEAGTLAYYEICDFLKGAEVHRLSNEKVPFATKGNQWVGYEHKES
                     VKNKVGFLKEKKLAGAMVWALDLDDFQGTCQPKEFFPLTNAIKDALA"
BASE COUNT      420 a    442 c    404 g    350 t
ORIGIN
```

Query Match 8.3%; Score 135; DB 81; Length 1616;
Best Local Similarity 62.5%; Pred. No. 3.23e-103;
Matches 674; Conservative 0; Mismatches 377; Indels 27; Gaps 13;

```
Db      88 caggctttgcggtcctgatgctgctccagagctgctctgcgtacaagctggtctgctact 147
Qy      24 CAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTGCTACT 83

Db     148 tcaccagctggtcccagtaccgggaaggcgttggaagcttcttaccagacgccatccaac 207
Qy      84 TCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACC 143

Db     208 ctttcctgtgcacccacatcatctacagctttgccaacatcagcagcgacaacatgctta 267
Qy     144 CCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCAC--CA-GCTGA 200

Db     268 gcacatgggagtggaatgacgagtcgaactatgacaagctgaataaactgaagaccagaa 327
Qy     201 GCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGATGA 260

Db     328 acaccaacctgaagaccctcctgtctgttggagggtggaaatttggcgaaaaaagatttt 387
Qy     261 ATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCGGCACTCAGAAGTTCA 320

Db     388 ccgagattgcctccaacactgagagacgcactgctttcgtccggtcggtagccccgttcc 447
Qy     321 CAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTTTC 380

Db     448 tgcgttcttatggctttgatgggctggatctcgcctggctctaccct------c------ 495
Qy     381 TGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGCCAGGGGA 440

Db     496 gct-taa-g-agacaagcagtatttctccaccctgatcaaggaactgaatgcggaattca 552
```

```
Qy     441 GCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAATGCCTTCC 500

Db     553 caaaggaggtccag--c-caggcagagagaaactcctgctcagcgcagctttgtcagcag 609
Qy     501 AGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCCAGCTG 560

Db     610 gaaaggtggccattgacactggctatgacatcgcccagatagcccaacacctggatttta 669
Qy     561 GGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATTTTG 620

Db     670 tcaatctcatgacctacgatttccatggagtctggcgccaaatcacaggccaccacagcc 729
Qy     621 TCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACATAACAGCC 680

Db     730 ccctcttccaaggccagaaggacactaggtttgacagatacagcaatgtgaactatgccg 789
Qy     681 CCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGATGCTGCTG 740

Db     790 tgcagtacatgatacgtctgggagcccaggccagcaagctactgatgggcatccccacct 849
Qy     741 TGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATGCCTACCT 800

Db     850 ttgggaagagcttcactctggcatcttc-tgaaa-atcagc-tgggagctccaatctcag 906
Qy     801 ACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCAGCCACAG 860

Db     907 gggaaggattaccaggccggttcaccaaggaggcagggaccctggcctactacgagatat 966
Qy     861 GGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTATGAAGTCT 920

Db     967 gcgacttcctcaaaggagctgaagtacatcgactctccaacgagaaggttcccttcgcta 1026
Qy     921 GCTCCTGG---AAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCCTACATCT 977

Db    1027 ccaagggcaaccagtgggtggggtatgagcacaaggagagtgtcaaaaacaaggttgggt 1086
Qy     978 TCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAGGTCAGCT 1037

Db    1087 tcctgaaggagaagaagctggcaggagccatggtgtgggcactggatttggatgattt 1144
Qy    1038 ATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGATGACTT 1095
```

```
RESULT   10
ID   MMD757     standard; RNA; ROD; 1506 BP.
AC   D87757;
NI   g1545818
DT   15-SEP-1996 (Rel. 49, Created)
DT   15-SEP-1996 (Rel. 49, Last updated, Version 1)
DE   House mouse; Musculus domesticus bone marrow cells mRNA for ECF-L
DE   precursor, complete cds.
KW   ECF-L precursor.
OS   Mus musculus (mouse)
OC   Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC   Theria; Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
RN   [1]
RP   1-1506
RA   Owhashi M.;
RT   ;
RL   Submitted (06-SEP-1996) to the EMBL/GenBank/DDBJ databases.
RL   Makoto Owhashi, The University of Tokushima, Faculty of Integrated
RL   Arts and Sciences; 1-1 Minami-Johsanjima, Tokushima, Tokushima 770,
RL   Japan (E-mail:ohashi@ias.tokushima-u.ac.jp, Tel:0886-56-7261,
RL   Fax:0886-56-7298)
RN   [2]
RP   1-1506
RA   Owhashi M., Arita H., Taoka Y., Hayai N.;
RT   "Molecular cloning and characterization of T lymphocyte-derived
RT   eosinophil chemotactic factor (ECF-L)";
RL   Unpublished.
FH   Key             Location/Qualifiers
FH
FT   source          1..1506
```

```
FT                   /organism="Mus musculus"
FT                   /cell_type="bone marrow cells"
FT                   /tissue_type="bone marrow"
FT   CDS             1..1197
FT                   /gene="mECF-L"
FT                   /codon_start=1
FT                   /product="ECF-L precursor"
FT                   /db_xref="PID:g1545819"
FT                   /translation="MAKLILVTGLAILLNVQLGSSYQLMCYYTSWAKDRPIEGSFKPG
  N
FT                   IDPCLCTHLIYAFAGMQNNEITYTHEQDLRDYEALNGLKDKNTELKTLLAIGGWKFGP
  A
FT                   SFSAMVSTPQNRQIFIQSVIRFLRQYNFDGLNLDWQYPGSRGSPPKDKHLFSVLVKEM
  R
FT                   KAFEEESVEKDIPRLLLTSTGAGIIDVIKSGYKIPELSQSLDYIQVMTYDLHDPKDGY
  T
FT                   GENSPLYKSPYDIGKSADLNVDSIISYWKDHGAASEKLIVGFPAYGHTFILSDPSKTG
  I
FT                   GAPTISTGPPGKYTDESGLLAYYEVCTFLNEGATEVWDAPQEVPYAYQGNEWVGYDNV
  R
FT                   SFKLKAQWLKDNNLGGAVVWRLDMDDFSGSFCHQRHFPLTSTLKGDLNIHSASCKGPY
  "
SQ   Sequence 1506 BP; 422 A; 337 C; 335 G; 412 T; 0 other;

  Query Match           7.5%;  Score 123;  DB 20;  Length 1506;
  Best Local Similarity  57.2%;  Pred. No. 5.92e-91;
  Matches   610;  Conservative    0;  Mismatches 451;  Indels   6;  Gaps   4;

Db     70 ctgatgtgctactataccagttgggctaaggacaggccaatagaagggagtttcaaacct 129
Qy     71 CTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCC 130

Db    130 ggtaatattgacccctgcctgtgtactcacctgatctatgcctttgctggaatgcagaat 189
Qy    131 AAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAAC 190

Db    190 aatgagatcacttacacacatgagcaagacttgcgtgactatgaagcattgaatggtctg 249
Qy    191 CACCAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTG 250

Db    250 aaagacaagaacactgagctaaaaactctcctggccattggaggatggaagtttggacct 309
Qy    251 AAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCGGCACT 310

Db    310 gcctcgttcagtgccatggtctctactcctcagaaccgtcagatattcattcagtcagtt 369
Qy    311 CAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCC 370

Db    370 atcagattccttcgtcaatataactttgatggcctcaacctggactggcagtaccctggg 429
Qy    371 ATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGA 430

Db    430 tctcgaggaagccctcctaaggacaaacatctcttcagtgttctggtgaaggaaatgcgt 489
Qy    431 AGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCC 490

Db    490 aaagcttttgaggaagaatctgtggagaaagacattccaaggctgctactcacttccaca 549
Qy    491 AATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCG 550

Db    550 ggagcaggaatcattgacgtaatcaagtctgggtacaagatccctgaactgtctcagtct 609
Qy    551 GTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAAC 610

Db    610 cttgactatattcaggtcatgacatatgatctccatgatcctaaggatggctacactgga 669
Qy    611 CTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGA 670

Db    670 gaaaatagtcccctctataaatctccatatgacattggaaagagtgctgatctcaatgtg 729
Qy    671 CATAACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTG 730

Db    730 gattcaatcatttcctactggaaggaccatggagcagcttctgagaagctcattgtggga 789
Qy    731 GATGCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGC 790

Db    790 tttccagcatatgggcataccttttatcctgagtgacccttctaagactggaattggtgcc 849
Qy    791 ATGCCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCC 850

Db    850 cctacaattagtactggcccaccaggaaagtacacagatgaatcaggactcctggcttac 909
Qy    851 CCAGCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTAC 910

Db    910 tatgaggtttgtacatttctgaatgaaggagccactgaggtctgggatgccccccaggaa 969
Qy    911 TATGAAGTCTG--C--TCCTGGAAGG-GG-GCCACCAAACAGAGAATCCAGGATCAGAAG 964

Db    970 gtaccctatgcctatcagggtaatgagtgggttggttatgacaatgtcaggagcttcaag 1029
Qy    965 GTGCCCTACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAA 1024

Db   1030 ttgaaggctcagtggctcaaggacaacaatttaggaggtgccgtggtctggcgcctggac 1089
Qy   1025 ACCAAGGTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGAC 1084

Db   1090 atggatgacttcagtggttctttctgtcaccagagacatttccctct 1136
Qy   1085 TTAGATGACTTTGCCGGCTTCTCCTGCAACCAGGGCCGATACCCCCT 1131


RESULT   11
LOCUS       D87757       1506 bp    mRNA            ROD       13-SEP-1996
DEFINITION  House mouse; Musculus domesticus bone marrow cells mRNA for ECF-L
            precursor, complete cds.
ACCESSION   D87757
NID         g1545818
KEYWORDS    ECF-L precursor.
SOURCE      Mus musculus bone marrow cells cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 1506)
  AUTHORS   Owhashi,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-1996) to the DDBJ/EMBL/GenBank databases. Makoto
            Owhashi, The University of Tokushima, Faculty of Integrated Arts
            and Sciences; 1-1 Minami-Johsanjima, Tokushima, Tokushima 770,
            Japan (E-mail:ohashi@ias.tokushima-u.ac.jp, Tel:0886-56-7261,
            Fax:0886-56-7298)
REFERENCE   2  (bases 1 to 1506)
  AUTHORS   Owhashi,M., Arita,H., Taoka,Y. and Hayai,N.
  TITLE     Molecular cloning and characterization of T lymphocyte-derived
            eosinophil chemotactic factor (ECF-L)
  JOURNAL   Unpublished (1996)
FEATURES             Location/Qualifiers
     source          1..1506
                     /organism="Mus musculus"
                     /cell_type="bone marrow cells"
                     /tissue_type="bone marrow"
     CDS             1..1197
                     /gene="mECF-L"
                     /codon_start=1
                     /product="ECF-L precursor"
                     /db_xref="PID:g1545819"
                     /translation="MAKLILVTGLAILLNVQLGSSYQLMCYYTSWAKDRPIEGSFKPG
                     NIDPCLCTHLIYAFAGMQNNEITYTHEQDLRDYEALNGLKDKNTELKTLLAIGGWKFG
                     PASFSAMVSTPQNRQIFIQSVIRFLRQYNFDGLNLDWQYPGSRGSPPKDKHLFSVLVK
                     EMRKAFEEESVEKDIPRLLLTSTGAGIIDVIKSGYKIPELSQSLDYIQVMTYDLHDPK
                     DGYTGENSPLYKSPYDIGKSADLNVDSIISYWKDHGAASEKLIVGFPAYGHTFILSDP
                     SKTGIGAPTISTGPPGKYTDESGLLAYYEVCTFLNEGATEVWDAPQEVPYAYQGNEWV
                     GYDNVRSFKLKAQWLKDNNLGGAVVWRLDMDDFSGSFCHQRHFPLTSTLKGDLNIHSA
                     SCKGPY"
BASE COUNT      422 a    337 c    335 g    412 t
ORIGIN
```

```
  Query Match             7.5%;  Score 123;  DB 81;  Length 1506;
  Best Local Similarity  57.2%;  Pred. No. 5.92e-91;
  Matches   610;  Conservative    0;  Mismatches 451;  Indels   6;  Gaps   4;

Db      70 ctgatgtgctactataccagttgggctaaggacaggccaatagaagggagtttcaaacct 129
           ||| | ||||||| |||| ||||| || |||| |     || |   | |||   ||
Qy      71 CTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCC 130

Db     130 ggtaatattgacccctgcctgtgtactcacctgatctatgcctttgctggaatgcagaat 189
             |  | |||||| |||| || || ||||| ||||| ||||| ||||| |||   ||
Qy     131 AAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAAC 190

Db     190 aatgagatcacttacacacatgagcaagacttgcgtgactatgaagcattgaatggtctg 249
             |  || | |    |||  |   | | |||  |  |  |||  | |  || ||||| |||
Qy     191 CACCAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTG 250

Db     250 aaagacaagaacactgagctaaaaactctcctggccattggaggatggaagtttggacct 309
           ||  | | |||  |  |||| || || ||  | ||||| ||||| ||||| || ||  ||
Qy     251 AAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCGGCACT 310

Db     310 gcctcgttcagtgccatggtctctactcctcagaaccgtcagatattcattcagtcagtt 369
             |||||  |  ||||| | ||  |  | ||||||||||  ||  |  | || |
Qy     311 CAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCC 370

Db     370 atcagattccttcgtcaatataactttgatggcctcaacctggactggcagtaccctggg 429
           ||||| || || ||  |||| | |||||| |||||  |||| |||||| ||||||| ||
Qy     371 ATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGA 430

Db     430 tctcgaggaagccctcctaaggacaaacatctcttcagtgttctggtgaaggaaatgcgt 489
              | || |||||| |    |||||  | | |||||     |||||  ||||  ||
Qy     431 AGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCC 490

Db     490 aaagcttttgaggaagaatctgtggagaaagacattccaaggctgctactcacttccaca 549
           || || ||  || | ||| |    |     ||  |    | | || || || | | |  |
Qy     491 AATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCG 550

Db     550 ggagcaggaatcattgacgtaatcaagtctgggtacaagatccctgaactgtctcagtct 609
           |   |||           |    |  |  |||| ||| || |     || |  | |||
Qy     551 GTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAAC 610

Db     610 cttgactatattcaggtcatgacatatgatctccatgatcctaaggatggctacactgga 669
           || || | | |   | ||| | || ||  ||||||   ||  |||     ||| |||
Qy     611 CTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGA 670

Db     670 gaaaatagtcccctctataaatctccatatgacattggaaagagtgctgatctcaatgtg 729
           | || || |||||||| ||    | | | || | |||       ||    ||||| |||
Qy     671 CATAACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTG 730

Db     730 gattcaatcatttcctactggaaggaccatggagcagcttctgagaagctcattgtggga 789
           ||| |     |     | |||  | |  | ||  |  || |    |||| ||  | ||
Qy     731 GATGCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGC 790

Db     790 tttccagcatatgggcataccttttatcctgagtgacccttctaagactggaattggtgcc 849
           | ||  | || || |   |||| |  |||    | | ||  | ||  || | || |||
Qy     791 ATGCCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCC 850

Db     850 cctacaattagtactggcccaccaggaaagtacacagatgaatcaggactcctggcttac 909
           ||  | |   |  ||||| | |||||    | |||  | |||  |||  | ||||| |||
Qy     851 CCAGCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTAC 910

Db     910 tatgaggtttgtacatttctgaatgaaggagccactgaggtctgggatgccccccaggaa 969
           ||||| || ||  |  | ||| | |  || |||||  |     |          ||| |
Qy     911 TATGAAGTCTG--C--TCCTGGAAGG-GG-GCCACCAAACAGAGAATCCAGGATCAGAAG 964

Db     970 gtaccctatgcctatcagggtaatgagtgggttggttatgacaatgtcaggagcttcaag 1029
           || |||||   ||  | ||  ||  ||||||| || | |||  ||||   |||||||||
Qy     965 GTGCCCTACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAA 1024

Db    1030 ttgaaggctcagtggctcaaggacaacaatttaggaggtgccgtggtctggcgcctggac 1089
             ||||     |  || ||| | ||     | || || ||| |||||||  ||||||
Qy    1025 ACCAAGGTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGAC 1084

Db    1090 atggatgacttcagtggttctttctgtcaccagagacatttccctct 1136
           | ||||||||    || |  | |||  ||||| | |  | ||| ||
Qy    1085 TTAGATGACTTTGCCGGCTTCTCCTGCAACCAGGGCCGATACCCCCT 1131


RESULT   12
LOCUS       SSU43490     2022 bp    mRNA            MAM       20-JAN-1996
DEFINITION  Sus scrofa oviductal secretory glycoprotein precursor mRNA,
            complete cds.
ACCESSION   U43490
NID         g1155331
KEYWORDS    .
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Artiodactyla; Suiformes; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 2022)
  AUTHORS   Buhi,W.C., Choi,I., Cleaver,B.D. and Simmen,F.A.
  TITLE     Molecular cloning and characterization of an estrogen-dependent
            porcine oviductal secretory glycoprotein (POSP)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2022)
  AUTHORS   Buhi,W.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-1995) William C. Buhi, University of Florida,
            Obstetrics and Gynecology, 1600 S.W. Archer Road, Gainesville, FL
            32610-0294, USA
FEATURES             Location/Qualifiers
     source          1..2022
                     /organism="Sus scrofa"
                     /strain="Crossbred Yorkshire x Duroc x Hampshire"
                     /clone="alv4"
                     /clone_lib="pig oviduct cDNA library of William C.Buhi"
                     /sex="female"
                     /cell_type="epithelial"
                     /tissue_type="oviduct"
                     /dev_stage="adult"
     5'UTR           <1..10
     CDS             11..1594
                     /codon_start=1
                     /product="oviductal secretory glycoprotein precursor"
                     /db_xref="PID:g1155332"
                     /translation="MGKLLLWVGLVLVLKHHNGAAHKLVCYFANWAFSRPGPASILPR
                     DLDPFLCTHLVFAFASMNDSQIVAKDARDESIFYPEFNQLKERNEKLKTLLSIGGWNF
                     GTSRFTTMLSTFTNREKFIRSAIGLLRTHGFDGLDLFFLYPGLRGSPRRDRWNFLFLL
                     EELLLAFRREAQLTMRPRLLLSAAVSADPHVIQKAYDVRLLGRLLDFINVLSYDLHGS
                     WEKVTGHNSPLFSLSDDPKSSAYTMNYWRKLGAPPEKLLMGFPTYGRTFRLLKASKNE
                     LGAEAVGPASPGKYTKQAGFLAYYEVCSFVQRAKKRWIDHQYVPYAYRGKEWVGYDDD
                     ISFSYKAFFIKKEHFGGAMVWTLDLDDVRGTFCGTGPFPLVYMLNDLLLKAEVSSTLS
                     PGFGLSTTVNSSRTCPESLAVTKDLTTDLGILPLGGEAVATETHGRSDNMTVTPGGGL
                     VAPTRPTLSFGKLTVAPEGKTESPGEKAMTPVGHPSVTPGDMSVPPVPIQTGDRITPP
                     RRQAVAPEKMTLPSGKRSD"
     sig_peptide     11..73
     mat_peptide     74..1591
                     /product="oviductal secretory glycoprotein"
     3'UTR           1595..2022
     polyA_signal    1987..1992
BASE COUNT      449 a    552 c    534 g    487 t
ORIGIN

  Query Match             6.4%;  Score 105;  DB 48;  Length 2022;
  Best Local Similarity  59.7%;  Pred. No. 8.60e-73;
  Matches   370;  Conservative    0;  Mismatches 247;  Indels   3;  Gaps   3;

Db      77 aagctggtgtgttattttgccaactgggcattcagtcggccaggccctgcttcgatcctg 136
           || ||||| || || ||  ||||||||||        | |  ||    |||    |||||
Qy      68 AAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTG 127

Db     137 ccccgggacctggacccctttctctgcacccacctggtcttcgccttcgcctcaatgaac 196
           |||  |||| ||||||||   || ||||||||||  ||| ||||||||    |||| |
Qy     128 CCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACC 187

Db     197 gacagtcagattgtggccaaggatgcccgggatgagagcatcttctacccagagttcaac 256
```

```
Qy    188 AACCACCAGCTGAGCACCACTGAGTGGAATGACGAGA-C-TCT-CTACCAGGAGTTCAAT 244

Db    257 cagctcaaggagaggaatgagaagctgaaaacactgctatccatcggagggtggaacttt 316

Qy    245 GGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTC 304

Db    317 ggcacatccaggttcaccacgatgctgtccaccttcaccaaccgggaaaagttcatcagg 376

Qy    305 GGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAAC 364

Db    377 tctgccataggcctcctgaggacacatggcttcgacgggctggacctcttcttcttgtac 436

Qy    365 TCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTAC 424

Db    437 cctggactcagaggcagccccaggcgtgaccgctggaactttctcttcttacttgaagag 496

Qy    425 CCAGGAAGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGAC 484

Db    497 ctcctgcttgccttcaggagggaggcacagctcaccatgcgtccgaggctgctgctctct 556

Qy    485 TTGGCCAATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGT 544

Db    557 gctgctgtctctgcggacccgcatgtcatccaaaaagcctatgatgtgcgccttctggga 616

Qy    545 GCAGCGGTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCC 604

Db    617 agactcctggatttcatcaatgtcttgtcttatgacttacatggaagctgggagaaggtc 676

Qy    605 CAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTC 664

Db    677 acgggacacaacagccccct 696

Qy    665 ACGGGACATAACAGCCCCCT 684
```

```
RESULT   13
LOCUS       OAU16719     2034 bp    mRNA            MAM       21-JUL-1995
DEFINITION  Ovis aries estrogen dependent oviduct protein precursor mRNA,
            complete cds.
ACCESSION   U16719
NID         g885600
KEYWORDS    .
SOURCE      sheep.
  ORGANISM  Ovis aries
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Caprinae; Ovis.
REFERENCE   1  (bases 1 to 2034)
  AUTHORS   DeSouza,M.M. and Murray,M.K.
  TITLE     An estrogen-dependent secretory protein, which shares identity with
            chitinases, is expressed in a temporally and regionally specific
            manner in the sheep oviduct at the time of fertilization and embryo
            development
  JOURNAL   Endocrinology 136 (6), 2485-2496 (1995)
  MEDLINE   95269691
REFERENCE   2  (bases 1 to 2034)
  AUTHORS   DeSouza,M.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1994) Mary M. DeSouza, Tufts University, Anatomy
            and Cell Biology, 136 Harrison Ave, Boston, MA 02111, USA
FEATURES             Location/Qualifiers
     source          1..2034
                     /organism="Ovis aries"
                     /clone="MD2A, MD34, MD42"
                     /clone_lib="sheep oviduct cDNA library of M. Murray"
                     /sex="female"
                     /tissue_type="oviduct"
     5'UTR           1..16
     sig_peptide     17..78
     CDS             17..1636
                     /codon_start=1
                     /product="estrogen dependent oviduct protein precursor"
                     /db_xref="PID:g885601"
                     /translation="MGKLLLWVGLLLMLKHHDGAAHKLVCYFTNWAFSRPGSASILPR
                     DLDPFLCTHLVFAFASMNNNQIVPKDPLDEKILYPEFNKLKERNRGLKTLLSVGGWNF
                     GTSRFTKMLSTFSNRERFVKSVIALLRTHGFDGLDLFFLYPGLRGSPARDRWTFVFLL
                     EELLQAFKNEAQLTMRPRLLLSAAVSGDPHVIQKAYDARLLGRLLDFISVLSYDLHGS
                     WEKVTGHNSPLFSLPGDPKSSAYAMSYWRQLGVPPEKLLMGLPTYGRTFHLLRASQNE
                     LGAGAAGPASPGKYTKQAGFLAYYEVCSFVQRAKKRWINDQYVPYAFKGKEWVGYDDA
                     ISFGYKAFFIKREHFGGAMVWTLDLDDFRGNFCGTGPFPLAHTLNNLLVNDEFSSTPS
                     PKFWFSTAVNSSRIGPEMPTMTRDLTTGLGILPLGGEAVATETHRKSATMTTTPRGET
                     ATPTRTPLSSGRRTAAPEGKTESPGEKPLTSVGRLAVSPGGIAVGPVHLQIGQKVTPP
                     GRKAGVPEKVTIPSGKMTVTPDGRAETLERL"
     mat_peptide     79..1629
                     /note="secretory glycoprotein, Mr 90-92,000, fimbria and
                     ampulla specific"
                     /product="estrogen dependent oviduct protein"
     3'UTR           1637..2034
     polyA_signal    1855..1860
     polyA_signal    2012..2017
BASE COUNT      451 a    548 c    537 g    498 t
ORIGIN
```

```
  Query Match             5.9%;  Score 97;  DB 47;  Length 2034;
  Best Local Similarity  59.0%;  Pred. No. 7.86e-65;
  Matches   366;  Conservative    0;  Mismatches 251;  Indels   3;  Gaps   2;

Db     83 aagctggtgtgttatttcaccaactgggcattcagtcggcccggctctgcctcgatcctg 142

Qy     68 AAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTG 127

Db    143 ccccgggatctggaccccttttctctgcacccacctggtatttgcctttgcctcgatgaac 202

Qy    128 CCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACC 187

Db    203 aacaatcagattgttcctaaggatcccctggatgagaaaatcctctacccagagttcaac 262

Qy    188 AACCACCAGCTGAGCACCACTGAGTGGAATGACGAGAC--TC-TCTACCAGGAGTTCAAT 244

Db    263 aagctcaaggagaggaacaggggctgaaaacactgctgtccgtcgggggtggaacttc 322

Qy    245 GGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTC 304

Db    323 ggcacctcgaggttcaccaagatgctgtccacgttttccaaccgggaaaggtttgtcaaa 382

Qy    305 GGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAAC 364

Db    383 tcagtgatcgccctcctgaggacacatggcttcgatggtctggacctcttcttcttgtac 442

Qy    365 TCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTAC 424

Db    443 cctggactcagaggcagccccgcgcgtgaccgctggacctttgtcttcttacttgaggag 502

Qy    425 CCAGGAAGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGAC 484

Db    503 ctcctgcaggccttcaagaatgaggcccagctcaccatgcgcccaaggctgctgctgtct 562

Qy    485 TTGGCCAATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGT 544

Db    563 gctgccgtctctggggacccccacgtcatccagaaagcatatgatgcacgccttctgggc 622

Qy    545 GCAGCGGTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCC 604

Db    623 agactcctggatttcatcagcgtcttgtcttacgacttacatggaagctgggaaaaggtc 682

Qy    605 CAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTC 664

Db    683 acaggacacaatagccctct 702

Qy    665 ACGGGACATAACAGCCCCCT 684
```

```
RESULT   14
LOCUS       OAU17988     1596 bp    mRNA            MAM       30-MAY-1996
DEFINITION  Ovis aries 92kD oestrus-associated oviduct glycoprotein mRNA,
```

```
            partial cds.
ACCESSION   U17988
NID         g602293
KEYWORDS    .
SOURCE      sheep.
  ORGANISM  Ovis aries
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Caprinae; Ovis.
REFERENCE   1  (bases 1 to 1596)
  AUTHORS   Marshall,J.T., Brownlee,A.G. and Nancarrow,C.D.
  TITLE     Cloning and sequencing of a cDNA encoding ovine oestrous
            glycoprotein
  JOURNAL   Reprod. Fert. Dev. 8 (2), 305-310 (1996)
REFERENCE   2  (bases 1 to 1596)
  AUTHORS   Marshall,J.T.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-1994) James T. A. Marshall, Division of Animal
            Production, Commonwealth Scientific and Industrial Research
            Organization, Locked Bag 1, Delivery Centre, Blacktown, New South
            Wales, 2148, Australia
FEATURES             Location/Qualifiers
     source          1..1596
                     /organism="Ovis aries"
                     /strain="merino"
                     /sex="female"
                     /cell_type="epithelial cell"
                     /tissue_type="ampulla oviduct"
                     /dev_stage="oestrus"
     CDS             <1..1596
                     /codon_start=1
                     /product="92kD oestrus-associated oviduct glycoprotein"
                     /db_xref="PID:g602294"
                     /translation="LLLVLKHHDGAAHKLVCYFTNWAFSRPGSASILPRDLDPFLCTH
                     LVFAFASMNNNQIVPKDPLDEKILYPEFNKLKERNRGLKTLLSVGGWNFGTSRFTKML
                     STFSNRERFVNSVIALLRTHGFDGLDLFFLYPGLRGSPARDRWTFVFLLEELLQAFKN
                     EAQLTMRPRLLLSAAVSGDPHVIQKAYDARLLGRLLDFISVLSYDLHGSWEKVTGHNS
                     PLFSLPGDPKSSAYAMSYWRQLGVPPEKLLMGLPTYGRTFHLLRASQNELGAGAVGPA
                     SPGKYTKQAGFLAYYEVCSFVQRAKKRWINDQYVPYAFKGKEWVGYDDAISFGYKAFF
                     IKREHFGGAMVWTLDLDDFRGNFCGTGPFPLVHTLNNLLVNDEFSSTPSPKFWFSTAV
                     NSSRIGPEMPTMTRDLTTGLGILPLGGEAVATETHRKSATMTTTPRGETATPTRTPLS
                     SGRRTAAPEGKTESPGEKPLTSVGHLAVSPGGIAVGPVHLQIGQKVTPPGRKAGVPEK
                     VTTPSGKMTVTPDGRAETLERRL"
     mRNA            <1..>1596
                     /product="92kD oestrus-associated oviduct glycoprotein"
     sig_peptide     <1..36
                     /note="inferred from bovine sequence: GenBank Accession
                     Number D16639"
                     /codon_start=1
     primer_bind     1..19
                     /note="derived from PCR primer based on bovine cDNA
                     sequence: GenBank Accession Number D16639"
     primer_bind     complement(1575..1596)
                     /note="derived from PCR primer based on bovine cDNA
                     sequence: GenBank Accession Number D16639"
BASE COUNT      352 a    440 c    430 g    374 t
ORIGIN

  Query Match           5.7%;  Score 93;  DB 47;  Length 1596;
  Best Local Similarity  58.7%;  Pred. No. 6.99e-61;
  Matches   364;  Conservative    0;  Mismatches 253;  Indels   3;  Gaps   2;

Db      40 aagctggtgtgttatttcaccaactgggcattcagtcggcccggctctgcctcgatcctg 99
           || ||||| || || ||||||||||||||      | |  ||    ||     |||||
Qy      68 AAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTG 127

Db     100 ccccgggatctggacccctttctctgcacccacctggtatttgcctttgcctcgatgaac 159
           |||  |||  ||||||||   || ||||||||||  | |  ||||| ||    |||| |
Qy     128 CCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACC 187

Db     160 aacaatcagattgttcctaaggatcccctggatgagaaaatcctctacccagagttcaac 219
           ||| | ||| |     | |  ||       || ||||   || ||||||  ||||||||
Qy     188 AACCACCAGCTGAGCACCACTGAGTGGAATGACGAGAC--TC-TCTACCAGGAGTTCAAT 244

Db     220 aagctcaaggagaggaacaggggctgaaaacactgctgtccgttgggggtggaacttc 279
           || ||| ||| |||      |||||| || ||| |  || | || || ||||| |||
Qy     245 GGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTC 304

Db     280 ggcacctcgaggttcaccaagatgctgtccacgttttccaaccgggaaaggtttgtcaat 339
           |||||   || ||||||  | ||| |  |||||     ||||||  | |  ||||||||
Qy     305 GGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAAC 364

Db     340 tcagtgatcgccctcctgaggacacatggcttcgatggtctggacctcttcttcttgtac 399
           || |  |||    | ||| | | |  |||| || || || ||||   ||    ||||
Qy     365 TCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTAC 424

Db     400 cctggactcagaggcagccccgcgcgtgaccgctggacctttgtcttcttacttgaggag 459
           || |||  |   || ||||| ||    |||     |  |||      | |  |  ||||
Qy     425 CCAGGAAGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGAC 484

Db     460 ctcctgcaggccttcaagaatgaggcccagctcaccatgcgcccaaggctgctgctgtct 519
           |     | |||||| || | || ||||||  | |   |     | | || || |||  |
Qy     485 TTGGCCAATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGT 544

Db     520 gctgccgtctctggggaccccacgtcatccagaaagcatatgatgcacgccttctgggc 579
           || || ||  | |  |  |    |    |  |    | ||| || |    |    | | |
Qy     545 GCAGCGGTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCC 604

Db     580 agactcctggatttcatcagcgtcttgtcttacgacttacacggaagctgggaaaaggtc 639
           |||||||||  ||| | |  || | |||||||| || ||    ||||| ||||||
Qy     605 CAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTC 664

Db     640 acaggacacaatagccctct 659
           || ||||| || ||||| ||
Qy     665 ACGGGACATAACAGCCCCCT 684


RESULT   15
LOCUS       BABEDOSG     2228 bp    mRNA            PRI       11-OCT-1996
DEFINITION  Papio hamadryas anubis estradiol-dependent oviduct-specific
            glycoprotein mRNA, partial cds.
ACCESSION   M59903
NID         g1256950
KEYWORDS    estradiol-dependent oviduct-specific glycoprotein.
SOURCE      Papio hamadryas anubis cDNA to mRNA.
  ORGANISM  Papio hamadryas anubis
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 2228)
  AUTHORS   Donnelly,K.M., Fazleabas,A.T., Verhage,H.G., Mavrogianis,P.A. and
            Jaffe,R.C.
  TITLE     Cloning of a recombinant complementary DNA to a baboon (Papio
            anubis) estradiol-dependent oviduct-specific glycoprotein
  JOURNAL   Mol. Endocrinol. 5 (3), 356-364 (1991)
  MEDLINE   91367180
REFERENCE   2  (bases 1 to 2228)
  AUTHORS   Jaffe,R.C., Arias,E.B., O'Day-Bowman,M.B., Donnelly,K.M.,
            Mavrogianis,P.A. and Verhage,H.G.
  TITLE     Regional Distribution and Hormonal Control of Estrogen-Dependent
            Oviduct Specific Glycoprotein mRNA in the Baboon (Papio anubis)
  JOURNAL   Biol. Reprod. (1996) In press
REFERENCE   3  (bases 1 to 2228)
  AUTHORS   Jaffe,R.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1991) Randal C. Jaffe, Department of Physiology
            and Biophysics, University of Illinois College of Medicine,
            Chicago, IL 60612-7342, USA
FEATURES             Location/Qualifiers
     source          1..2228
                     /organism="Papio hamadryas anubis"
     CDS             13..1884
                     /codon_start=1
                     /product="estradiol-dependent oviduct-specific
                     glycoprotein"
                     /db_xref="PID:g1256951"
```

```
                     /translation="MWKLLLWVGLVLVLKHHDGAAHKLVCYFTNWAHSRPGPASILPH
                     DLDPFLCTHLIFAFASMNNNQIVAKDLQDEKILYPEFNKLKERNRELKTLLSIGGWNF
                     GTSRFTTMLSTFANREKFIASVISLLRTHDFDGLDLFFLYPGLRGSPMHDRWTFLFLI
                     EELLFAFRKEALLTMRPRLLLSAAVSGVPHIVQTSYDVRFLGRLLDFINVLSYDLHGS
                     WEKFTGHNSPLFSLPEDPKSSAYAMNYWRKLGAPSEKLIMGIPTYGRTFRLLKASKNG
                     LQATAIGPASPGKYTKQAGFLAYFEICSFVWGAKKHWIDYQYVPYANKGKEWVGYDDA
                     ISFSYKAWFIRREHFGGAMVWTLDMDDVRGTFCGTGPFPLVYVMNDILVRAEFSSTSL
                     PQFWLSSAVNSSSTDPERLAVTKAWTTDIKILPPGGEAGVTEIHGKCENMTITPRVTI
                     VTPTKETVSLGKHTVALGEKTEITGATTMTSVGHQSMTPGEKALTPVGHQSELPGKKT
                     LTPVGHQSVTTGQKTLISVGYHSVTPGEKTLTPVGHPSVTPVSHQSVSPGGMTMTPVH
                     FQTETLRQNTMAPRRKAVAHEKVTVPSRKISVTPEGQTVPLRGEYLTSETGTHPQDG"
BASE COUNT      559 a    572 c    531 g    566 t
ORIGIN

  Query Match           5.7%;  Score 93;  DB 67;  Length 2228;
  Best Local Similarity  58.7%;  Pred. No. 6.99e-61;
  Matches   364;  Conservative    0;  Mismatches 253;  Indels   3;  Gaps   1;

Db     79 aaactggtgtgttatttcaccaactgggcacacagtcggccaggccccgcctcgatcttg 138
Qy     68 AAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTG 127

Db    139 ccccatgacctggaccccttttctctgcacccacctgatatttgcctttgcctcaatgaac 198
Qy    128 CCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACC 187

Db    199 aacaatcagattgttgctaaggatctccaggacgagaaaattctctacccagagttcaac 258
Qy    188 AACCACCAGCTGAGCACCACTGAGTGGAATGACGAGAC---TCTCTACCAGGAGTTCAAT 244

Db    259 aaactaaaggagaggaacagagagctgaaaacactgctgtccatcggcgggtggaacttt 318
Qy    245 GGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTC 304

Db    319 ggcacctccaggttcaccactatgttgtccacatttgccaaccgtgaaaagtttattgct 378
Qy    305 GGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAAC 364

Db    379 tcagttatatcccttctgaggacacatgactttgatggtcttgatctttttttcttatat 438
Qy    365 TCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTAC 424

Db    439 cctggactaagaggcagccccatgcatgaccggtggacttttctcttcttaattgaagag 498
Qy    425 CCAGGAAGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGAC 484

Db    499 ctcctgtttgccttccggaaggaggcactgctcaccatgcgcccaaggctgctgctgtct 558
Qy    485 TTGGCCAATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGT 544

Db    559 gctgctgtttctggggtcccacacatcgtccaaacatcctatgatgtgcgctttctagga 618
Qy    545 GCAGCGGTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCC 604

Db    619 agactcctggatttcatcaatgtcttgtcttatgacttacatggaagttgggaaaagttc 678
Qy    605 CAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTC 664

Db    679 acaggacataatagcccccct 698
Qy    665 ACGGGACATAACAGCCCCCT 684


Search completed: Sun Jun 29 19:55:25 1997
Job time : 1542 secs.
```

****

****

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Jun 29 19:55:44 1997; MasPar time 175.81 Seconds
860.863 Million cell updates/sec

Tabular output not generated.

Title: >US-08-663-618A-1
Description: (1-1636) from US08663618A.seq
Perfect Score: 1636
N.A. Sequence: 1 CATGGTGCGGTCTGTGGCCT.....CCCTCTTCCCAAAAAAAAAA 1636
Comp: GTACCACGCCAGACACCGGA.....GGGAGAAGGGTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 121476 seqs, 46255616 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq26
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23

Statistics: Mean 9.561; Variance 5.543; scale 1.725

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	178	10.9	1681	13	Q85245	YKL-40 gene.	7.87e-103
	2	133	8.1	966	1	N81756	Gene encoding polypep	4.48e-71
	3	93	5.7	2504	15	Q90443	Murine oviduct specif	1.67e-43
	4	92	5.6	1047	2	Q10572	Human Natriuretic Pep	7.92e-43
	5	85	5.2	1994	15	Q90442	Bovine oviduct specif	4.11e-38
c	6	78	4.8	1047	2	Q10572	Human Natriuretic Pep	1.92e-33
	7	69	4.2	2366	15	Q90444	Hamster oviduct speci	1.61e-27
	8	51	3.1	204	1	N81164	Base substituted E.co	4.79e-16
c	9	45	2.8	91	9	Q51746	Oligonucleotide probe	2.21e-12
	10	42	2.6	91	9	Q51746	Oligonucleotide probe	1.37e-10
	11	38	2.3	114	12	Q70468	Generic DNA sequence	2.95e-08
	12	37	2.3	114	12	Q70467	Generic DNA sequence	1.11e-07
	13	37	2.3	114	12	Q70469	Generic DNA sequence	1.11e-07
c	14	38	2.3	204	1	N81164	Base substituted E.co	2.95e-08
	15	36	2.2	114	12	Q70465	Generic DNA sequence	4.09e-07
	16	36	2.2	114	12	Q70466	Generic DNA sequence	4.09e-07
	17	35	2.1	114	12	Q70473	Generic DNA sequence	1.50e-06
	18	34	2.1	114	12	Q70470	Generic DNA sequence	5.42e-06
	19	34	2.1	114	12	Q70472	Generic DNA sequence	5.42e-06
	20	34	2.1	114	12	Q70471	Generic DNA sequence	5.42e-06
c	21	34	2.1	114	12	Q70469	Generic DNA sequence	5.42e-06
c	22	33	2.0	114	12	Q70466	Generic DNA sequence	1.93e-05
c	23	33	2.0	114	12	Q70467	Generic DNA sequence	1.93e-05
c	24	32	2.0	114	12	Q70465	Generic DNA sequence	6.81e-05
c	25	32	2.0	114	12	Q70468	Generic DNA sequence	6.81e-05
c	26	31	1.9	114	12	Q70470	Generic DNA sequence	2.37e-04
c	27	31	1.9	114	12	Q70473	Generic DNA sequence	2.37e-04
	28	31	1.9	1167	6	Q37704	Aphanocladium album m	2.37e-04
	29	31	1.9	1167	11	Q68353	Aphanocladium album m	2.37e-04
	30	31	1.9	1320	6	Q37714	Aphanocladium album m	2.37e-04
	31	31	1.9	1364	6	Q37713	Aphanocladium album m	2.37e-04
	32	31	1.9	1405	6	Q37711	Aphanocladium album p	2.37e-04
	33	31	1.9	1701	6	Q37712	Aphanocladium album p	2.37e-04
	34	30	1.8	1074	12	Q78145	Trichoderma harzianum	8.09e-04
	35	30	1.8	1554	12	Q78144	Trichoderma harzianum	8.09e-04
c	36	27	1.7	39	7	Q51787	Mixed oligonucleotide	2.92e-02
c	37	28	1.7	114	12	Q70471	Generic DNA sequence	9.01e-03
c	38	27	1.7	114	12	Q70472	Generic DNA sequence	2.92e-02
	39	27	1.7	3871	2	N71302	HSV-1 gB and surround	2.92e-02
	40	27	1.7	6180	23	T36387	Periplasmic chitodext	2.92e-02
c	41	26	1.6	565	6	Q35072	HCV envelope region n	9.31e-02
c	42	25	1.5	82	21	T13610	DC43 TSAR library gen	2.90e-01
	43	25	1.5	146	21	T13624	ME#2b library generat	2.90e-01
	44	25	1.5	498	3	N50034	Sequence encoding new	2.90e-01
	45	25	1.5	2951	23	T36390	Soluble chitinase cod	2.90e-01

ALIGNMENTS

RESULT 1
ID Q85245 standard; cDNA; 1681 BP.
AC Q85245;
DT 12-AUG-1995 (first entry)
DE YKL-40 gene.
KW YKL-40; diagnosis; prognosis; therapy; breast cancer; metastasis;
KW marker; joint disease; connective tissue; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 135..1681
FT /*tag= a
PN WO9501995-A.
PD 19-JAN-1995.
PF 08-JUL-1994; U07754.
PR 09-JUL-1993; US-089989.
PA (REGC) UNIV CALIFORNIA.
PI Johansen JS, Price PA;
DR WPI; 95-066866/09.
PT Use of YKL-40 and anti-YKL-40 antibodies - for developing prods.
PT for diagnosis prognosis and therapy of diseases involving
PT connective tissue degradation.
PS Disclosure; Page 66-67; 88pp; English.
CC YKL-40 (40 kDa) was purified from human osteosarcoma MG63 cells.
CC The N-terminal sequence is shown in R70745; the full coding region
CC of the YKL-40 gene is given in Q85245. Homology of the N-terminal
CC and 2 internal peptides (R70746-47) with a bacterial polysaccharide
CC hydrolase suggests that YKL-40 degrades polysaccharide components
CC of connective tissue. YKL-40 is a marker of e.g. metastatic breast
CC cancer and inflammatory or degenerative joint diseases.
SQ Sequence 1681 BP; 401 A; 503 C; 421 G; 356 T;

Query Match 10.9%; Score 178; DB 13; Length 1681;
Best Local Similarity 63.6%; Pred. No. 7.87e-103;
Matches 698; Conservative 0; Mismatches 376; Indels 24; Gaps 10;

```
Db     94 caggctttgtggtcctggtgctgctccagtgctgctctgcatacaaactggtctgctact 153
          |||| ||  |||||||| || || |||  ||  |||||||    ||||||||||||||||
Qy     24 CAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTGCTACT 83

Db    154 acaccagctggtcccagtaccgggaaggcgatgggagctgcttcccagatgccccttgacc 213
          ||||| |||| |||||||| |  | || || |   ||| | | ||  | | | | ||||
```

```
Qy      84 TCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACC 143

Db     214 gcttcctgtgtacccacatcatctacagctttgccaatataagcaacgatcacatcgaca 273
Qy     144 CCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCTGAGCA 203

Db     274 cctgggagtggaatgatgtgacgctctacggcatgctcaacacactcaacaacacgaacc 333
Qy     204 CCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGATGAATC 263

Db     334 ccaacctgaagactctcttgtctgtcggaggatggaactttgggtctcaaagattttcca 393
Qy     264 CCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCGGCACTCAGAAGTTCACAG 323

Db     394 agatagcctccaacacccagagtcgccggactttcatcaagtcagtaccgccatttctgc 453
Qy     324 ATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTTTCTGC 383

Db     454 gcacccatggctttgatgggcgtgaccttgcctggctctaccctgga---c---ggag-- 505
Qy     384 GCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGCCAGGGGAGCC 443

Db     506 -------agacaaacaccatttttaccaccctaatcaaggaaatgaaggccgaatttataa 558
Qy     444 CTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAATGCCTTCCAGC 503

Db     559 aggaagcccag--c-cagggaaaaagcagctcctgctcagcgcagcactgtctgcgggga 615
Qy     504 AGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCCAGCTGGGC 563

Db     616 aggtcaccattgacagcagctatgacattgccaagatatcccaacacctggatttcatta 675
Qy     564 AGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATTTTGTCA 623

Db     676 gcatcatgacctacgattttcatggcgcctggcgtgggaccacaggccatcacagtcccc 735
Qy     624 ACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACATAACAGCCCCC 683

Db     736 tgttccgaggtcaggaggatgcaagtcctgacagattcagcaacactgactatgctgtgg 795
Qy     684 TCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGATGCTGCTGTGC 743

Db     796 ggtacatgttgaggctgggggctcctgccagtaagctggtgatgggcatccccaccttcg 855
Qy     744 AACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATGCCTACCTACG 803

Db     856 ggaggagcttcactctggcttcttc-tgagactg--gtgttccagcgccaatctcaggac 912
Qy     804 GACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCAGCCACAGGGT 863

Db     913 cgggaattccaggccggttcaccaaggaggcagggacccttgcctactatgagatctgtg 972
Qy     864 CTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTATGAAGTCTGCT 923

Db     973 acttcctccgcggagccacagtccatagaaccctcggccagcaggtcccctatgccacca 1032
Qy     924 CCTGGAAG-G-GG-GCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCCTACATCTTCC 980

Db    1033 agggcaaccagtgggtaggatacgacgaccaggaaagcgtcaaaagcaaggtgcagtacc 1092
Qy     981 GGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAGGTCAGCTATC 1040

Db    1093 tgaaggataggcagctggcaggcgccatggtatgggccctggacctggatgacttccagg 1152
Qy    1041 TGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGATGACTTTGCCG 1100

Db    1153 gctccttctgcggccagg 1170
Qy    1101 GCTTCTCCTGCAACCAGG 1118


RESULT    2
ID   N81756 standard; DNA; 966 BP.
AC   N81756;
DT   19-OCT-1990  (first entry)
DE   Gene encoding polypeptide involved in protective mechanisms
KW   Immune response; cell growth; ss.
FH   Key             Location/Qualifiers
FT   CDS             1..966
FT   /*tag=  a
PN   J63032898-A.
PD   30-JAN-1988.
PF   16-JUL-1986; 167518.
PR   16-JUL-1986; JP-167518.
PA   (DAIN) Dainippon Pharm KK.
DR   WPI; 88-068419/10.
DR   P-PSDB; P81342.
PT   New polypeptide and DNA encoding it - related to protective mechanisms
PT   such as immune response etc.
PS   Disclosure; p; Japanese.
CC   Polypeptide encoded is involved in protective mechanisms such as immune
CC   response, cell growth and activation of protective functions.
SQ   Sequence   966 BP;   238 A;   282 C;   247 G;   199 T;

  Query Match             8.1%;  Score 133;  DB 1;  Length 966;
  Best Local Similarity  66.3%;  Pred. No. 4.48e-71;
  Matches   270;  Conservative    0;  Mismatches 137;  Indels   0;  Gaps   0;

Db     23 caggctttgtggtcctggtgctgctccagtgctgctctgcatacaaactggtctgctact 82
Qy     24 CAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTGCTACT 83

Db     83 acaccagctggtcccagtaccgggaaggcgatgggagctgcttcccagatgcccttgacc 142
Qy     84 TCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACC 143

Db    143 gcttcctctgtacccacatcatctacagctttgccaatataagcaacgatcacatcgaca 202
Qy    144 CCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCTGAGCA 203

Db    203 cctgggagtggaatgatgtgacgctctacggcatgctcaacacactcaagaacaggaacc 262
Qy    204 CCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGATGAATC 263

Db    263 ccaacctgaagactctcttgtctgtcggaggatggaactttgggtctcaaagattttcca 322
Qy    264 CCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCGGCACTCAGAAGTTCACAG 323

Db    323 agatagcctccaacacccagagtcgccggactttcatcaagtcagtaccgccatttctgc 382
Qy    324 ATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTTTCTGC 383

Db    383 gcacccatggctttgatggggtggaccttgcctggctctaccctgga 429
Qy    384 GCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGA 430


RESULT    3
ID   Q90443 standard; cDNA to mRNA; 2504 BP.
AC   Q90443;
DT   10-JAN-1996  (first entry)
DE   Murine oviduct specific glycoprotein cDNA.
KW   Murine oviduct specific glycoprotein; recombinant production;
KW   MOGP; ds.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   CDS             3..2160
FT   /*tag=  a
FT   /note= "START codon absent"
FT   sig_peptide     3..56
FT   /*tag=  b
FT   mat_peptide     57..2156
FT   /*tag=  c
PN   J07107979-A.
PD   25-APR-1995.
PF   15-AUG-1994; 214227.
```

```
PR   19-AUG-1993; JP-227881.
PA   (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR   WPI; 95-190179/25.
DR   P-PSDB; R90443.
PT   New DNA encoding an oviduct-specific glycoprotein - useful for
PT   recombinant protein production in high quantities.
PS   Claim 1; Pages 11-14; 22pp; Japanese.
CC   Q90443 encodes R73992 murine oviduct specific glycoprotein (MOGP).
CC   The cDNA can be used for the commercial recombinant prodn. of
CC   MOGP in high quantities.
SQ   Sequence   2504 BP;    678 A;    616 C;    538 G;    672 T;

  Query Match           5.7%;  Score 93;  DB 15;  Length 2504;
  Best Local Similarity  58.7%;  Pred. No. 1.67e-43;
  Matches   364;  Conservative    0;  Mismatches 253;  Indels   3;  Gaps   2;

Db      60 aaactggtgtgctatttcaccaattgggcacatagtcggccaggccctgcctccatcatg 119
           ||||||||  |||||  |||||||| ||||| ||     | |  ||    ||   | || ||
Qy      68 AAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTG 127

Db     120 ccccatgacctggaccccttttctttgtacgcatctgatatttgcctttgcctcaatgagc 179
           |||  | |||  ||||||||    |||||  || || || || |  |||||  ||    |||| |
Qy     128 CCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACC 187

Db     180 aacaatcagattgtggccaagaatctgcaggatgaaaacgttctctatccagagttcaac 239
           |||  | |||  | |  |  ||  | |  |  | ||||  |   ||||||  |  ||||||||
Qy     188 AACCACCAGCT-GAG--CACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAAT 244

Db     240 aaactcaaggaaaggaacagggagctgaaaacactgctgtccattggagggtggaacttc 299
              || |||  | |  |||     |||||||  || |||  |  ||||  |||||  ||||| |||
Qy     245 GGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTC 304

Db     300 ggcacatcacggttcaccgctatgctctccacacttgccaaccgggaaaagtttattgac 359
           |||||      ||||||  |  ||||  |  ||||      ||||||  |  |  |||  |  ||
Qy     305 GGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAAC 364

Db     360 tcagtcatatccttcctgagaatacatggttttgatggtcttgacctttttcttcttgtac 419
           ||  |  |||    ||  |||  |  |  |  |  |||||  ||  |||||||||  ||    ||||
Qy     365 TCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTAC 424

Db     420 cctggactacgaggcagtcccccacacgaccggtggaatttttctcttcttaattgaagag 479
           ||  |||    |  ||  ||  ||  |    |||  |  |  |  ||      |  |  |  ||
Qy     425 CCAGGAAGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGAC 484

Db     480 ctccagtttgccttcgagagggaggccctgctcactcagcacccacggctgctgctgtcg 539
             |      |||||||  ||  |||  ||||  |  |  |    |  |    |||  ||  ||  |||
Qy     485 TTGGCCAATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGT 544

Db     540 gctgctgtctctggcatcccaagcatcatccatacgtcttatgatgcgctccttttagga 599
           ||  ||  ||  |  |       |  |  |     |  ||  |     ||  ||  |  |  |     |  |
Qy     545 GCAGCGGTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCC 604

Db     600 aggcgtttggatttcattaatgttttgtcttatgacttacatggaagttgggaaaagttt 659
             |    |||||||  | ||  || || | || ||||| |||||   ||||||  ||| |
Qy     605 CAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTC 664

Db     660 acaggacacaatagtcctct 679
           ||  |||||  ||  ||  ||  ||
Qy     665 ACGGGACATAACAGCCCCCT 684


RESULT    4
ID   Q10572 standard; DNA; 1047 BP.
AC   Q10572;
DT   09-APR-1991  (first entry)
DE   Human Natriuretic Peptide Receptor B.
KW   NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW   hyperaldosteronism; glaucoma; guanyl cyclase.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide          1..22
FT   /label= signal sequence
FT   Protein          12
FT   /label= mature NPBR
FT   Domain           23..455
FT   /label= extracellular domain
FT   /note= "binds natriuretic peptides A,B and C]"
FT   Domain           456..456
FT   /label= transmembrane domain
FT   Domain           479..1047
FT   /label= cytoplasmic domain
FT   /note= "GC and protien kinase activity"
FT   Modified -site   24..26
FT   /label= N-glycos_site
FT   Modified -site   35..37
FT   /label= N-glycos_site
FT   Modified -site   161..163
FT   /label= N-glycos_site
FT   Modified -site   195..197
FT   /label= N-glycos_site
FT   Modified -site   244..246
FT   /label= N-glycos_site
FT   Modified -site   277..279
FT   /label= N-glycos_site
FT   Modified -site   349..351
FT   /label= N-glycos_site
FT   Modified -site   600..602
FT   /label= N-glycos_site
PN   WO9100292-A.
PD   10-JAN-1991.
PF   22-JUN-1990; U03586.
PR   23-JUN-1989; US-370673.
PA   (GETH ) GENENTECH INC.
PI   Chang M, Goeddel D, Lowe D;
DR   WPI; 91-036711/05.
DR   N-PSDB; Q10324.
PT   Natriuretic protein receptor B - for diagnosis and treatment of
PT   kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS   Claim 3; Fig 1; 49pp; English.
CC   The sequence was derived from the DNA encoding natriuretic peptide
CC   receptor B, NPRB, having guanyl cyclase (GC) activity and protein
CC   kinase activity.  The DNA can be inserted into expression vectors
CC   for the prodn. of the protein, opt. after being mutated to produce
CC   NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr-
CC   114,952).  The protein (or variants) can be used in treatment of
CC   natriuretic peptide disorders, and also to isolate peptides using
CC   affinity chromatography.  Antibodies with affinity for NPRB can
CC   also be prepd.
SQ   Sequence   1047 BP;    87 A;    15 C;    83 G;    51 T;

  Query Match           5.6%;  Score 92;  DB 2;  Length 1047;
  Best Local Similarity  10.6%;  Pred. No. 7.92e-43;
  Matches   106;  Conservative  279;  Mismatches 603;  Indels  15;  Gaps  14;

Db       5 snnnnvaanaggvrnngarnntnavvnnnhnnsyawawnrvgnavanavnangrannvdn 64
           :    : | || ::  | :     ::   :  ::  |: ::   :   :   |     :
Qy      95 GCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACCCCAGCCTTTGC 154

Db      65 rnvssnnngacsnynannsavdnknyhdndnnngngcvynaas-varnashwrnnnntag 123
           : ::     | :        :: :  :: :      |:: | : :|: |:::     | |
Qy     155 ACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCTGAGCACCACTGAGTGG 214

Db     124 avasgnsaknd hyrtnvrtgnsanknggnnvvtnhghnnwtaraannyndartddrnhynt 183
           |: :  :|: ::: | : |  :| : |  ::   |:    |:||  : :|    :  :
Qy     215 AATGACGAG-ACTCT-C-TACCAGGAGTTCAATGGCCTGAAGAAGATGAATCCCAAGCTG 271

Db     184 nngvnnanngsnnsvnhnvyarnnggnnnathnnrangrnvyncgnnnmnhnnnnnan-r 242
             |:       :: : : |:   |    |:  :     :   |   : :        :
Qy     272 AAGACCCTGTTAGCCATCGGAGGCTGGAATTTCGGCACTCAGAAGTTCACAGATATGGTA 331

Db     243 nnntngdyvnnyndvngnsnragntratgrnwndnrtrnnnananrnanntvnvntyrnnn 302
             |:::  : :: | : :|  |      : : :      | :    |: :   :
Qy     332 GCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTTTCTGCGCAAATAC 391

Db     303 nnnynnnnnrnnnrarndngvnngnsnmnnnagcnydgnnnyanvnnntnnnggtrndgn 362
             :    :    :        :     | |  :|   :  :        |:  |
```

```
Qy     392 AGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGCCAGGGGAGCCC-TGCCGT 450

Db     363 rnvnkmngrryhgvtgnvvmdknndrntdnvnwamgdndsgdnnnaahysganknnwwtg 422
           : :  : |::: |:|  ::::   :: |: :  |: : :: :     ::: |     : |
Qy     451 AGACAAGGAGC-GCTTCACAACCCTGGTACAGGACTTGGCCAATGCCTTCCAGCAGGAAG 509

Db     423 rnnn-wv-kgannsdnnncandnddnscdktnnstnanvangtgntnnmngvssnnnnrk 480
             :: : |  ::    | :  : : : |  :|   :  ||     : |:::     :
Qy     510 CCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCCAGCTGGGCAGACCT 569

Db     481 nmnnknnasmnwrnrwnnnnngnsnryhkgagsrntnsnrgssygsnmtahgkynnnant 540
           :   :: :: ::     | : :::  | ::     : :::|:  | :|::     |
Qy     570 ATGTGGATGC-TGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATTTTGTCAACCTT 628

Db     541 ghnkgnvvankhvnkkrnnntrnvnnnnkhmrdvnnnhntrnngacndnnnncnvtnycn 600
            : :| : | ::: :      :    :: :::   :  :   |  :    | :  :
Qy     629 -ATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACATAACAGCCCCCTCTA 687

Db     601 rgsnndnnnndsnnndwmnrysnnndnvkgmannhnsnnsshgsnkssncvvdsrnvnkn 660
           :   :    ::  :     : :: :|  : :  :::|: :::  : : : :
Qy     688 CAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGATGCTGCTGTGCAACA 747

Db     661 tdygnasnrstannddnnanyakknntannnnsgnnnnttgmnaadvysngnnnnnnanr 720
           : |  : :: |  ::    : :          |    |||  | :::          | :
Qy     748 GTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATGCCTACCTACGGACG 807

Db     721 sgnnynngndnsnknnvnkvrngnrnynrnsndrtnnnnnnvnnmnrcwandnanrndng 780
           :  :   :  :          :  : : ::      :  : :| | :     : |
Qy     808 CTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCAGCCACAGGGTCTGG 867

Db     781 nnkgnnrrnnknggtsnndnnnnrmnnyannnnknvnnrtnaynnnkrkananynnnnh 840
              ::       |:       ::       :       |:   ::      :
Qy     868 CACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTATGAAGTCTGCTCCTG 927

Db     841 svannnkrgntvnanandsvtnynsdnvgntansanstnmnvvtnnndnytcnd--annd 898
           ::|   ::   : | | :::    :  :| |  :| :    ::|   : :|| :  |  :
Qy     928 GAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCCTACATCTTCCGGGACAA 987

Db     899 nndvykvntngdaymvvsgnngrngnrhannnarmananndavssnrnrhrnhdnnrnrn 958
             ::::: | |: :  :     : | : |     :| |  :|::  :  :: ::  :
Qy     988 CCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAGGTCAGCTATCTGAAGCA 1047

Db     959 gvhtgnvcagvvgnkmnrycnngdtvntasrmnsngnanknhv 1001
           |:: | :| |::  :  ::|  |:|: | ::: : | | : ::
Qy    1048 GAAGGGACTGGGCGG-GGCCATGGTC-TGGGCACTGGACTTAG 1088


RESULT    5
ID   Q90442 standard; cDNA to mRNA; 1994 BP.
AC   Q90442;
DT   10-JAN-1996  (first entry)
DE   Bovine oviduct specific glycoprotein cDNA.
KW   Bovine oviduct specific glycoprotein; recombinant production;
KW   BOGP; ds.
OS   Bos taurus.
FH   Key             Location/Qualifiers
FT   CDS             2..1615
FT   /*tag= a
FT   /note= "START codon absent"
FT   sig_peptide     2..55
FT   /*tag= b
FT   mat_peptide     56..1612
FT   /*tag= c
PN   J07107979-A.
PD   25-APR-1995.
PF   15-AUG-1994; 214227.
PR   19-AUG-1993; JP-227881.
PA   (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR   WPI; 95-190179/25.
DR   P-PSDB; R73991.
PT   New DNA encoding an oviduct-specific glycoprotein - useful for
PT   recombinant protein production in high quantities.
PS   Claim 1; Pages 7-9; 22pp; Japanese.

CC   Q90442 encodes R73991 bovine oviduct specific glycoprotein (BOGP).
CC   The cDNA can be used for the commercial recombinant prodn. of
CC   BOGP in high quantities.
SQ   Sequence   1994 BP;    445 A;    536 C;    520 G;    493 T;

  Query Match             5.2%;  Score 85;  DB 15;  Length 1994;
  Best Local Similarity  58.1%;  Pred. No. 4.11e-38;
  Matches   360;  Conservative    0;  Mismatches 257;  Indels   3;  Gaps   2;

Db      59 aagctggtgtgttatttcaccaactgggcattcagtcggcccggccctgcctcaatcctg 118
           || ||||| || || ||||||||||||||        | |  ||    ||     |||||
Qy      68 AAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTG 127

Db     119 cctcgggatctggacccgtttctctgcacccacctggtatttgcctttgcctcgatgagc 178
           ||   |||  |||||||    || ||||||||||  | |  ||||| ||    |||| |
Qy     128 CCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACC 187

Db     179 aacaatcagattgttcctaaggatccccaggatgagaaaatcctctacccagagttcaac 238
           ||| | ||| |   | |  ||     | || ||||   || ||||||  ||||||||
Qy     188 AACCACCAGCTGAGCACCACTGAGTGGAATGACGAGAC--TC-TCTACCAGGAGTTCAAT 244

Db     239 aagctcaaggagaggaacagagggctgaaaacgctgctgtccatcgggggtggaacttt 298
           || ||| ||| |||      |||||| || ||| |  ||||||| || ||||| ||
Qy     245 GGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTC 304

Db     299 ggcaccgtgaggttcaccacgatgctgtccacgttttccaaccgggaaaggttcgtcagt 358
           |||||   || ||||||    ||| |  |||||     ||||||  | |  || ||||
Qy     305 GGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAAC 364

Db     359 tcagtgatcgccctcctgaggacacatggctttgatggtctggacctcttcttcttgtac 418
           || |  |||    | ||| | | | |  ||||||| || || |||||   ||    ||||
Qy     365 TCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTAC 424

Db     419 cctggactcagaggcagccctgcgcgtgaccgctggacctttgtcttcttacttgaagag 478
           || |||  |   || ||||||||    |||      |  |||     | | |  | ||
Qy     425 CCAGGAAGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGAC 484

Db     479 ctcctgcaggccttcaagaatgaggcccagctcaccatgcgtccaaggctgctgctgtct 538
           |     | |||||| || | || ||||||  | |   |     | | || || |||  |
Qy     485 TTGGCCAATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGT 544

Db     539 gctgccgtctctggggacccccatgtcgtccagaaagcgtatgaagcacgccttctgggc 598
           || || ||  | |  | |         ||  |    |  || || |    |    | | |
Qy     545 GCAGCGGTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCC 604

Db     599 agactcctggatttcatcagtgtcttgtcttatgacttacatggaagctgggaaaaggtc 658
                |||||||||  |||   |  || | || ||||| ||||    ||||| ||||||
Qy     605 CAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTC 664

Db     659 acaggacacaatagccctct 678
           || ||||| || ||||| ||
Qy     665 ACGGGACATAACAGCCCCCT 684


RESULT    6
ID   Q10572 standard; DNA; 1047 BP.
AC   Q10572;
DT   09-APR-1991  (first entry)
DE   Human Natriuretic Peptide Receptor B.
KW   NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW   hyperaldosteronism; glaucoma; guanyl cyclase.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide          1..22
FT   /label= signal sequence
FT   Protein          12
FT   /label= mature NPBR
FT   Domain           23..455
FT   /label= extracellular domain
FT   /note= "binds natriuretic peptides A,B and C]"
FT   Domain           456..456
FT   /label= transmembrane domain
FT   Domain           479..1047
```

```
FT   /label= cytoplasmic domain
FT   /note= "GC and protien kinase activity"
FT   Modified -site  24..26
FT   /label= N-glycos_site
FT   Modified -site  35..37
FT   /label= N-glycos_site
FT   Modified -site  161..163
FT   /label= N-glycos_site
FT   Modified -site  195..197
FT   /label= N-glycos_site
FT   Modified -site  244..246
FT   /label= N-glycos_site
FT   Modified -site  277..279
FT   /label= N-glycos_site
FT   Modified -site  349..351
FT   /label= N-glycos_site
FT   Modified -site  600..602
FT   /label= N-glycos_site
PN   WO9100292-A.
PD   10-JAN-1991.
PF   22-JUN-1990; U03586.
PR   23-JUN-1989; US-370673.
PA   (GETH ) GENENTECH INC.
PI   Chang M, Goeddel D, Lowe D;
DR   WPI; 91-036711/05.
DR   N-PSDB; Q10324.
PT   Natriuretic protein receptor B - for diagnosis and treatment of
PT   kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS   Claim 3; Fig 1; 49pp; English.
CC   The sequence was derived from the DNA encoding natriuretic peptide
CC   receptor B, NPRB, having guanyl cyclase (GC) activity and protein
CC   kinase activity.  The DNA can be inserted into expression vectors
CC   for the prodn. of the protein, opt. after being mutated to produce
CC   NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC   114,952).   The protein (or variants) can be used in treatment of
CC   natriuretic peptide disorders, and also to isolate peptides using
CC   affinity chromatography.  Antibodies with affinity for NPRB can
CC   also be prepd.
SQ   Sequence   1047 BP;    87 A;    15 C;    83 G;    51 T;

  Query Match            4.8%;  Score 78;  DB 2;  Length 1047;
  Best Local Similarity   8.9%;  Pred. No. 1.92e-33;
  Matches    83;  Conservative  251;  Mismatches 596;  Indels   7;  Gaps   7;

Db     28 avvnnnhnnsyawawnrvgnavanavnangrannvdnrnvssnnngacsnynannsavdn 87
Cp   1623 AAGAGGGGCACAAACCAAAGATTTATTTTGCAAGTGAAAGGGGCAGCCCAGGAGACCCAG 1564

Db     88 knyhdndnnngngcvynaasvarnashwrnnnntagavasgnsakndhyrtnvrt-gnsa 146
Cp   1563 AAAAAAGGAAGGCAAGGCTGAGAGCAGAAAGCCTGGATAAAGGAAGACCACAGAAAGGCC 1504

Db    147 nkngnnvvtnhghnnwtaraannyndartddrnhyntnngvnnanngsnnsvnhnvyarn 206
Cp   1503 TGCAGGAGCCAGATTGCGGCCCCCAGGGAAAACCCAGGAGGCAGGCTGTAGAGTGATCCT 1444

Db    207 nggnnnathnnrangrnvyncgnnnmnhnnnnnanrnnntngdyvnnyndvngnsnragn 266
Cp   1443 GGGCCCAGCCTCAA-AGCTGGGACTGGAGGGGCTTTAGCGACTCAATTCCAGGTGCAGCA 1385

Db    267 tratgrnwndnrtrnnananrnanntvnvntyrnnnnnnynnnnnrnnnrarndngvnng 326
Cp   1384 TT-TGCAGGAGTTGCTGAACACCAGGCCTGTCGGGCAGCTTTGCTGGAACAGCCGCCCCG 1326

Db    327 nsnmnnnagcnydgnnnyanvnnntnnnggtrndgnrnvnkmngrryhgvtgnvvmdknn 386
Cp   1325 CTGCACAGCTGTAGAAGCTGGACCGTTCCCGAGGATTGGGATAGAGCCCATCAGCTTTGC 1266

Db    387 drntdnvnwamgdndsgdnnnaahysganknnwwtgrnnnwvkgannsdnnncandnddn 446
Cp   1265 CCTGGCAGAACGTGTCTTGTCCAGGGCTGGGGCCATGCTCAGGTTCAGAGGGCTGACCTG 1206

Db    447 scdktnnstnanvangtgntnnmngvssnnnnrknmnnknnasmnwrnrwnnnnngnsnr 506
Cp   1205 GTTTTGGAACTTCAAGCTCTGGGGTGCCTGAAGGCAAGTATGGAAGACTCAGTTCCTGCC 1146

Db    507 yhkgagsrntnsnrgssygsnmtahgkynnnantghnkgnvvankhvnkkrnnntrnvnn 566
Cp   1145 GTAGCGTCTGGATGAGGGGGTATCGGCCCTGGTTGCAGGAGAAGCCGGCAAAGTCATCTA 1086

Db    567 nnkhmrdvnnnhntrnngacndnnnncnvtnycnrgsnndnnnndsnnndwmnrysnnnd 626
Cp   1085 AGTCCAGTGCCCAGACCATGGCCCCGCCCAGTCCCTTCTGCTTCAGATAGCTGACCTTGG 1026

Db    627 nvkgmannhnsnnsshgsnkssncvvdsrnvnkntdygnasnrstannddnnanyakknn 686
Cp   1025 TTTTGAAGCTCTCCACATCATCAAAGCCCACCCACTGGTTGTCCCGGAAGAT-GTAGGGC 967

Db    687 tannnnsgnnnnttgmnaadvysngnnnnnnanrsgnnynngndnsnknnvnkvrngnrn 746
Cp   966 ACCTTCTGATCCTGGATTCTCTGTTTGGTGGCCCCCTTCCAGGAGCAGACTTCATAGTAG 907

Db    747 -ynrnsndrtnnnnnnvnnmnrcwandnanrndngnnkgnnrrnnknggtsnndnnnnrm 805
Cp   906 GCCAGCATCCCTCCTTCCTTGGTGAAGGGGCCTGGAGTGCCAGACCCTGTGGCTGGGGCC 847

Db    806 nnyannnnknvnnrtnaynnnkrkanannynnnnhsvannnkrgntvnanandsvtnyns 865
Cp   846 CCCACTCTGGTGTCTGATGAGGAGGCCAGTGTGAAGGAGCGTCCGTAGGTAGGCATGC-C 788

Db    866 dnvgntansanstnmnvvtnnndnytcndanndnndvykvntngdaymvvsgnngrngnr 925
Cp   787 AAGGATCAGCT-TGCTGGCAGGGGTCCCCTTCTGCAGCCACTGTTGCACAGCAGCATCCA 729

Db    926 hannnarmananndavssnrnrhrnhdnnrnrngvht 962
Cp   728 CGTTGAGGCTGGCTGCTGCACCACTCTCTTCTTGCCT 692


RESULT    7
ID   Q90444 standard; cDNA to mRNA; 2366 BP.
AC   Q90444;
DT   10-JAN-1996  (first entry)
DE   Hamster oviduct specific glycoprotein cDNA.
KW   Hamster oviduct specific glycoprotein; recombinant production;
KW   HOGP; ds.
OS   Cricetulus griseus.
FH   Key             Location/Qualifiers
FT   CDS             3..2009
FT   /*tag= a
FT   /note= "START codon absent"
FT   sig_peptide     3..56
FT   /*tag= b
FT   mat_peptide     57..2006
FT   /*tag= c
PN   J07107979-A.
PD   25-APR-1995.
PF   15-AUG-1994; 214227.
PR   19-AUG-1993; JP-227881.
PA   (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR   WPI; 95-190179/25.
DR   P-PSDB; R73993.
PT   New DNA encoding an oviduct-specific glycoprotein - useful for
PT   recombinant protein production in high quantities.
PS   Claim 1; Pages 16-18; 22pp; Japanese.
CC   Q90444 encodes R73993 hamster oviduct specific glycoprotein (HOGP).
CC   The cDNA can be used for the commercial recombinant prodn. of
CC   HOGP in high quantities.
SQ   Sequence   2366 BP;    602 A;    574 C;    573 G;    617 T;

  Query Match            4.2%;  Score 69;  DB 15;  Length 2366;
  Best Local Similarity  61.1%;  Pred. No. 1.61e-27;
  Matches   232;  Conservative    0;  Mismatches 145;  Indels   3;  Gaps   2;

Db     60 aagctggtctgctatttcaccaactgggctcacagtcggccagtccctgcctccatcttg 119
```

```
Qy      68 AAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTG 127

Db     120 ccccgtgacctggatccctttctttgtacacacctgatatttgcctttgcctcgatgagc 179
           |||   ||| ||||  |||   |||||  || |||||  ||  |  |||||  ||    ||||  |
Qy     128 CCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACC 187

Db     180 aacaatcagattgttgccaataatctccaggatgagaaaattctctatccagagttcaac 239
           |||  |  |||  |     |||  |  |   |   |  ||||   |  |  ||||||  |  ||||||||
Qy     188 AACCACCAGCTGAGCACCACTGAG-TGGA--ATGACGAGACTCTCTACCAGGAGTTCAAT 244

Db     240 aaactcaaggagaggaacagagccctgaaaacactactgtctgttggaggctggaacttc 299
             ||  |||  |||  |||       |||||  ||  ||   |   |   |  ||||||||||| |||
Qy     245 GGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTC 304

Db     300 ggcacatcacggttcaccactatgctgtccacccttgccagccgtgaaaaatttattggt 359
           |||||      ||||||   ||||  |  ||||      ||  ||||  |  |  |||  |
Qy     305 GGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAAC 364

Db     360 tcagttgtatccttcctgagaacacatggctttgatgggcttgatctcttcttcttgtac 419
           ||  |   |    ||  |||  |  |  |  |||||||  ||  |||||  ||   ||    ||||
Qy     365 TCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTAC 424

Db     420 cctggactacgaggcagccc 439
           ||  |||   |  ||  |||||
Qy     425 CCAGGAAGCCAGGGGAGCCC 444


RESULT    8
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT   /*tag=  a
FT   /function=multiple cloning site
FT   primer_bind     187..204
FT   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match             3.1%;  Score 51;  DB 1;  Length 204;
  Best Local Similarity  15.4%;  Pred. No. 4.79e-16;
  Matches    18;  Conservative   62;  Mismatches  36;  Indels   1;  Gaps   1;

Db      75 yycdchvgccgymrttthhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhdd 134
           ::|: :: | | : ||| ::::: : :::::     | ::||  ::::::    | ::::
Qy     603 CCCAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGG 662

Db     135 hyvybbbvynvhnhnncncccbnnhvchnvhbnnhrnwayvrhdarrddvhccvchc 191
           :::::::::  : :  | |||:  ::|: :::  :: :| :::: :::::: |: :|
Qy     663 TCACGGGACATAACAGCCCCCTC-TACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCC 718


RESULT    9
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match             2.8%;  Score 45;  DB 9;  Length 91;
  Best Local Similarity   2.0%;  Pred. No. 2.21e-12;
  Matches     1;  Conservative   47;  Mismatches   3;  Indels   0;  Gaps   0;

Db      10 gssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsv 60
           |:::::::::: :::::: :::::::::::::::::::::::::::: :::::::
Cp    1361 GGCCTGTCGGGCAGCTTTGCTGGAACAGCCGCCCCGCTGCACAGCTGTAGA 1311


RESULT   10
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match             2.6%;  Score 42;  DB 9;  Length 91;
  Best Local Similarity   3.8%;  Pred. No. 1.37e-10;
  Matches     2;  Conservative   45;  Mismatches   5;  Indels   0;  Gaps   0;

Db      12 svhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsvctc 63
           :::::::::: :::::::::::::   :::::::::::::::::::: ||
Qy     124 CCTGCCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTC 175
```

```
RESULT   11
ID   Q70468 standard; DNA; 114 BP.
AC   Q70468;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT   /*tag= a
FT   /note= "this sequence represents 'Z'; Z can be a
FT   sequence of 6, 9 or 12 nucleotides (see
FT   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65154.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed activity
CC   allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match             2.3%;  Score 38;  DB 12;  Length 114;
  Best Local Similarity   4.5%;  Pred. No. 2.95e-08;
  Matches     5;  Conservative   34;  Mismatches  73;  Indels   0;  Gaps   0;

Db       3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnn 62
           :  :  :  :  :  :  :  :  :  :  :|||  :  :  :  :  :  :
Qy     121 CTTCCTGCCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGG 180

Db      63 bnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
           :  :  :  :  :  :  : ||  :  :  :  :  :  :  :  :  :  :
Qy     181 CATGACCAACCACCAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTAC 232


RESULT   12
ID   Q70467 standard; DNA; 114 BP.
AC   Q70467;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT   /*tag= a
FT   /note= "this sequence represents 'Z'; Z can be a
FT   sequence of 6, 9 or 12 nucleotides (see
FT   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65153.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)16(TGC)(NNB)1Z(NNB)16(TGC)(NNB)1Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need for
CC   complex methods of hybridoma formation or in vivo antibody production.
CC   The TSARs are easily characterised and have designed activity allowing
CC   direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match             2.3%;  Score 37;  DB 12;  Length 114;
  Best Local Similarity   4.8%;  Pred. No. 1.11e-07;
  Matches     5;  Conservative   32;  Mismatches  68;  Indels   0;  Gaps   0;

Db       6 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnnnnnnnnbnn 65
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|||  :        :
Qy     109 GGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTA 168

Db      66 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtg 110
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :||
Qy     169 CGCCTTCGCTGGCATGACCAACCACCAGCTGAGCACCACTGAGTG 213


RESULT   13
ID   Q70469 standard; DNA; 114 BP.
AC   Q70469;
DT   07-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT   /*tag= a
```

```
FT   /note= "this sequence represents 'Z'; Z can be a
FT   sequence of 6,9 or 12 nucleotides (see
FT   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC   This generic formula can be represented  as follows: X(TGC)(NNB)10-
CC   (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking  restriction
CC   sites (X is not the same as Y) that are not specified further. This
CC   sequence generates peptides that are cloverleaf in structure. Other
CC   generic sequences are shown in Q70465-68. Other specific peptides
CC   generated by these generic sequences are shown in R65150-54. TSARs are
CC   concatenated heterofunctional proteins or peptides, comprising at least
CC   two functional regions - a binding domain with affinity for a ligand and
CC   a second effector peptide portion that is chemically or biologically
CC   active.They may further comprise a linker peptide between the 2 domains.
CC   The oligonucleotides are also designed so that the expressed peptide
CC   contains 2 or 4 cysteine residues positioned in, or flanking, the
CC   unpredicted or variant residues. These residues confer some degree of
CC   conformational rigidity to the peptides. The TSARs or compsns. comprising
CC   a TSAR binding domain can be used in vivo to deliver a chemically or
CC   biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC   or enzyme, to the specific target or on the cell. They can also replace
CC   the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC   and therefore circumvent the need for complex methods of hybridoma
CC   formation or in vivo antibody production. The TSARs are easily
CC   characterised and have designed activity allowing direct and rapid
CC   detection in a screening process.
SQ   Sequence   114 BP;    0 A;    4 C;    4 G;    4 T;

  Query Match             2.3%;  Score 37;  DB 12;  Length 114;
  Best Local Similarity   6.3%;  Pred. No. 1.11e-07;
  Matches     7;  Conservative   32;  Mismatches  72;  Indels   0;  Gaps   0;

Db      2 gcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnn 61
          ||  :  :  :  :  :  :  :  :  :|||  :  :  :  :  :  :
Qy    120 GCTTCCTGCCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTG 179

Db     62 nbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbt 112
            :  :  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|
Qy    180 GCATGACCAACCACCAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCT 230


RESULT   14
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT   /*tag= a
FT   /function=multiple cloning site
FT   primer_bind     187..204
FT   /*tag= b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match             2.3%;  Score 38;  DB 1;  Length 204;
  Best Local Similarity  11.5%;  Pred. No. 2.95e-08;
  Matches    11;  Conservative   49;  Mismatches  36;  Indels   0;  Gaps   0;

Db     91 thhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddhyvybbbvynvhnhnn 150
          |:::::: : :  :  :::| |: || :  :  :| :   : : : ::::::  :: :
Cp   1125 TATCGGCCCTGGTTGCAGGAGAAGCCGGCAAAGTCATCTAAGTCCAGTGCCCAGACCATG 1066

Db    151 cncccbnnhvchnvhbnnhrnwayvrhdarrddvhc 186
            |||:  :: : :::  :: : :::::|: ::::|
Cp   1065 GCCCCGCCCAGTCCCTTCTGCTTCAGATAGCTGACC 1030


RESULT   15
ID   Q70465 standard; DNA; 114 BP.
AC   Q70465;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT   /*tag= a
FT   /note= "this sequence represents 'Z'; Z can be a
FT   sequence of 6, 9 or 12 nucleotides (see
FT   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65150 and R65151.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
```

CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compsns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARs are easily characterised and have designed
CC activity allowing direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 2.2%; Score 36; DB 12; Length 114;
Best Local Similarity 3.6%; Pred. No. 4.09e-07;
Matches 4; Conservative 34; Mismatches 74; Indels 0; Gaps 0;

```
Db      3 bnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :  :  :  :  : ||  :  :  :  :  :  :  :  :  :  :  :  :
Qy    130 CAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAA 189

Db     63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnb 114
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :| |  :  :  :
Qy    190 CCACCAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTC 241
```

Search completed: Sun Jun 29 19:58:47 1997
Job time : 183 secs.

****

(TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 27 16:52:08 1997; MasPar time 11.66 Seconds
435.449 Million cell updates/sec

Tabular output not generated.

Title: >US-08-663-618A-2
Description: (1-466) from US08663618A.pep
Perfect Score: 3443
Sequence: 1 MVRSVAWAGFMVLLMIPWGS.....QQSCPTGLVFSNSCKCCTWN 466

Scoring table: PAM 150
Gap 11

Searched: 92623 seqs, 10896596 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq26
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19

Statistics: Mean 36.077; Variance 158.164; scale 0.228

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1350	39.2	668	14	R73993	Hamster oviduct speci	2.13e-115
2	1345	39.1	537	14	R73991	Bovine oviduct specif	6.48e-115
3	1332	38.7	718	14	R73992	Murine oviduct specif	1.17e-113
4	912	26.5	321	1	P81342	Polypeptide involved	3.09e-73
5	597	17.3	371	13	R70025	Tobacco chitinase enc	2.36e-43
6	597	17.3	377	13	R70029	Tobacco cluster-A pro	2.36e-43
7	581	16.9	424	12	R63441	Trichoderma harzianum	7.47e-42
8	576	16.7	389	10	R56861	Aphanocladium album m	2.20e-41
9	576	16.7	423	6	R33072	Aphanocladium album p	2.20e-41
10	569	16.5	389	6	R33068	Aphanocladium album m	9.94e-41
11	463	13.4	866	19	W02159	Soluble chitinase.	7.19e-31
12	389	11.3	561	12	R64823	Chitinase A.	4.43e-24
13	372	10.8	561	9	R48981	Bacterial chitinase-A	1.56e-22
14	271	7.9	238	6	R32547	HaNPV IE-1 gene prote	1.65e-13
15	219	6.4	1046	19	W02156	Periplasmic chitodext	5.16e-09
16	139	4.0	25	13	R70751	YKL-40 N-terminal pep	1.79e-02
17	139	4.0	25	12	R70745	YKL-40 N-terminal seq	1.79e-02
18	136	4.0	29	11	R60073	Bovine Oviduct Glycop	3.04e-02
19	99	2.9	398	18	R95856	S. pyogenes speB gene	1.58e+01
20	99	2.9	428	6	R31372	Murine bone morphogen	1.58e+01
21	99	2.9	428	16	R86905	Murine BMP-9.	1.58e+01
22	99	2.9	553	14	R80445	Murine soluble epoxid	1.58e+01
23	101	2.9	958	2	R07575	Glucoamylase encoded	1.15e+01
24	101	2.9	958	1	P81181	Sequence of glucoamyl	1.15e+01
25	101	2.9	1070	4	R21521	Alpha galactosidase.	1.15e+01
26	95	2.8	438	9	R47259	Pre-pro-VGR1.	2.97e+01
27	92	2.7	1356	5	R26999	Novel type III RTK en	4.73e+01
28	91	2.6	19	13	R70752	YKL-40 internal pepti	5.51e+01
29	91	2.6	19	12	R70746	YKL-40 internal pepti	5.51e+01
30	90	2.6	256	17	R97235	MuTu putative oncogen	6.43e+01
31	90	2.6	459	17	R88058	Protein encdoed by Mu	6.43e+01
32	88	2.6	513	9	R47264	Pre-pro BMP6.	8.71e+01
33	88	2.6	513	2	R07311	Human Bone Morphogene	8.71e+01
34	88	2.6	513	6	R32904	Human BMP-6 peptide s	8.71e+01
35	88	2.6	513	7	R36735	Human BMP-6.	8.71e+01
36	90	2.6	575	3	P50018	Sequence encoded by t	6.43e+01
37	90	2.6	709	2	P70423	FixD nitrogen fixatio	6.43e+01
38	90	2.6	709	3	P50017	Sequence encoded by t	6.43e+01
39	88	2.6	896	17	R92750	Human EGF receptor su	8.71e+01
40	87	2.5	281	18	R94503	1-acylglycerol-3-phos	1.01e+02
41	87	2.5	281	16	R87724	Full length meadowfoa	1.01e+02
42	87	2.5	284	19	R99248	Limnanthes 2-acyltran	1.01e+02
43	87	2.5	369	18	R98125	Human lymphocyte cell	1.01e+02
44	87	2.5	542	16	R90772	Bacillus stearothermo	1.01e+02
45	87	2.5	2351	3	P50319	Human antihaemophilic	1.01e+02

ALIGNMENTS

RESULT 1
ID R73993 standard; Protein; 668 AA.
AC R73993;
DT 10-JAN-1996 (first entry)
DE Hamster oviduct specific glycoprotein.
KW Hamster oviduct specific glycoprotein; recombinant production;
KW HOGP.
OS Cricetulus griseus.
FH Key Location/Qualifiers
FT Peptide 1..18
FT /label= sig_peptide
FT Peptide 19..668
FT /label= mat_peptide
PN J07107979-A.
PD 25-APR-1995.
PF 15-AUG-1994; 214227.
PR 19-AUG-1993; JP-227881.
PA (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR WPI; 95-190179/25.
DR N-PSDB; Q90444.
PT New DNA encoding an oviduct-specific glycoprotein - useful for
PT recombinant protein production in high quantities.
PS Claim 4; Pages 16-18; 22pp; Japanese.
CC Q90444 encodes R73993 hamster oviduct specific glycoprotein (HOGP).
CC The cDNA can be used for the commercial recombinant prodn. of
CC HOGP in high quantities.
SQ Sequence 668 AA;

Query Match 39.2%; Score 1350; DB 14; Length 668;
Best Local Similarity 48.6%; Pred. No. 2.13e-115;
Matches 186; Conservative 83; Mismatches 103; Indels 11; Gaps 9;

```
Db      4 wvg-lvllmkpndgtayklvcyftnwahsrpvpasilprdldpflcthlifafasmsnnq 62
            | | :|||| |  |:| ||||||||||: |   | :||:|||| ||||||:|||:|:|:|
Qy      7 WAGFMVLLMIPW-GSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQ 65

Db     63 ivannlqdekilypefnklkernralktllsvggwnfgtsrfttmlstlasrekfigsvv 122
           : : :  ||  || ||| ||  |  |||||::|||||| :|| |::|  :|: |: | :
Qy     66 LSTTEWNDET-LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAI 124

Db    123 sflrthgfdgldlfflypglrgspindrwnflflieelqfafekealltqrprlllsaav 182
            |||  :||||||   ||| :|||  |:  |  |:::|  ||: ||  : : ||||||||
Qy    125 RFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAV 184
```

```
Db    183 -sgipyiiqtsydvhllgrrldfinvlsydlhgswekstghnsplfslpedpkssa-fam 240
          :|  |: :::|:|  ::: |||:|:::||:|||||| |||||||:   |:: ::| : :
Qy    185 PAGQTYV-DAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNV 243

Db    241 n--y--wrnlgapadkllmgfpaygrtfhllreskng1qaasmgpaspgkytkqagflay 296
          :    |   |:|| ||::| |:|||:| |   |   : |:: |:::|| :||::| |||
Qy    244 DAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAY 303

Db    297 yevcsfiqraekhwidhqyvpyaykgkewvgyddavsfsykamfvkkehfggamvwtldm 356
          |||||    | |::|: | ||| :: ::|||:||  ||  |  ::|   :||||||:||:
Qy    304 YEVCSW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDL 362

Db    357 ddvrgtfcgngpfplvhilnell 379
          ||  |  |  | :||:: | : |
Qy    363 DDFAGFSCNQGRYPLIQTLRQEL 385


RESULT    2
ID    R73991 standard; Protein; 537 AA.
AC    R73991;
DT    10-JAN-1996  (first entry)
DE    Bovine oviduct specific glycoprotein.
KW    Bovine oviduct specific glycoprotein; recombinant production;
KW    BOGP.
OS    Bos taurus.
FH    Key             Location/Qualifiers
FT    Peptide         1..18
FT    /label= sig_peptide
FT    Peptide         19..537
FT    /label= mat_peptide
PN    J07107979-A.
PD    25-APR-1995.
PF    15-AUG-1994; 214227.
PR    19-AUG-1993; JP-227881.
PA    (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR    WPI; 95-190179/25.
DR    N-PSDB; Q90442.
PT    New DNA encoding an oviduct-specific glycoprotein - useful for
PT    recombinant protein production in high quantities.
PS    Claim 4; Pages 7-9; 22pp; Japanese.
CC    Q90442 encodes R73991 bovine oviduct specific glycoprotein (BOGP).
CC    The cDNA can be used for the commercial recombinant prodn. of
CC    BOGP in high quantities.
SQ    Sequence   537 AA;

  Query Match           39.1%;  Score 1345;  DB 14;  Length 537;
  Best Local Similarity  50.4%;  Pred. No. 6.48e-115;
  Matches   191;  Conservative   80;  Mismatches 101;  Indels   7;  Gaps   6;

Db      6 glllvlkhhdgaahklvcyftnwafsrpgpasilprdldpflcthlvfafasmsnnqivp 65
          |::::|:   |:| |||||||||||  | | | :||:|||| ||||||::|||:|:|:|:
Qy      9 GFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLST 68

Db     66 kdpqdekilypefnklkernrglktllsiggwnfgtvrfttmlstfsnrerfvssviall 125
          :  ||  || ||| ||  |  |||||:|||||||| :|| |::| :||: ||:| | :|
Qy     69 TEWNDET-LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFL 127

Db    126 rthgfdgldlfflypglrgspardrwtfvflleellqafkneaqltmrprlllsaavsgd 185
          |  :||||||   ||| :|||| |:  |  |:::|  ||  ||| : : |||||||::
Qy    128 RKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG 187

Db    186 phvvqkayearllgrlldfisvlsydlhgswekvtghnsplfs-l--pgdpks-sa-yam 240
            |: :||   ::: |||:::::||:|||||||||||||:     :| : | :   |:
Qy    188 QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAV 247

Db    241 nywrqlgvppekllmglptygrtfhllkasqnelraqavgpaspgkytkqagflayyeic 300
            | | | |: ||::|:|||||:| |  :|:  : | | |:::|| :||::| |||||:|
Qy    248 QQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVC 307

Db    301 cfvrrakkrwindqyvpyafkgkewvgyddaisfgykaffikrehfggamvwtldlddfr 360
            : | |::| || ||| |: ::|||:||  ||  |  ::|:  :||||||:||||||
Qy    308 SW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFA 366

Db    361 gyfcgtgpfplvhtlnnll 379
          |: |  | :||::||   |
Qy    367 GFSCNQGRYPLIQTLRQEL 385


RESULT    3
ID    R73992 standard; Protein; 718 AA.
AC    R73992;
DT    10-JAN-1996  (first entry)
DE    Murine oviduct specific glycoprotein.
KW    Murine oviduct specific glycoprotein; recombinant production;
KW    MOGP.
OS    Mus musculus.
FH    Key             Location/Qualifiers
FT    Peptide         1..18
FT    /label= sig_peptide
FT    Peptide         19..718
FT    /label= mat_peptide
PN    J07107979-A.
PD    25-APR-1995.
PF    15-AUG-1994; 214227.
PR    19-AUG-1993; JP-227881.
PA    (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR    WPI; 95-190179/25.
DR    N-PSDB; Q90443.
PT    New DNA encoding an oviduct-specific glycoprotein - useful for
PT    recombinant protein production in high quantities.
PS    Claim 4; Pages 11-14; 22pp; Japanese.
CC    Q90443 encodes R73992 murine oviduct specific glycoprotein (MOGP).
CC    The cDNA can be used for the commercial recombinant prodn. of
CC    MOGP in high quantities.
SQ    Sequence   718 AA;

  Query Match           38.7%;  Score 1332;  DB 14;  Length 718;
  Best Local Similarity  49.2%;  Pred. No. 1.17e-113;
  Matches   191;  Conservative   81;  Mismatches 107;  Indels   9;  Gaps   6;

Db      7 lvllmkhsdgtayklvcyftnwahsrpgpasimphdldpflcthlifafasmsnnqivak 66
          :|||||    |:| ||||||||||: | | | ::| |||| ||||||:|||:|:|:|: :
Qy     11 MVLLMIPW-GSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTT 69

Db     67 nlqdenvlypefnklkernrelktllsiggwnfgtsrftamlstlanrekfidsvisflr 126
          :  ||  || ||| ||  |  |||||:|||||||| :|| |::|  ||: |::| | |||
Qy     70 EWNDET-LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLR 128

Db    127 ihgfdgldlfflypglrgspphdrwnflflieelqfaferealltqhprlllsaavsgip 186
            :||||||   ||| :|||: |:  |  |:::|  ||::||  :   ||||||||::
Qy    129 KYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQ 188

Db    187 siihtsydalllgrrldfinvlsydlhgswekftghnsplfslpedskssa--y---amn 241
          : : ::|:   ::: |||:|:::||:|||||| |||||||:   |:| ::|      |:
Qy    189 TYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQ 248

Db    242 ywrklgtpadklimgfptygrnfyllkesknglqtasmgpaspgkytkqagflayyevcs 301
          |   |||| |||:| |||||:| |   |   : ::: |:::|| :||::| ||||||||
Qy    249 QWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCS 308

Db    302 fvqrakkhwidyqyvpyafkgkewlgyddtisfsykamyvkrehfggamvwtldmddvrg 361
              | |::|: | ||| |: ::|:|:||  ||  |  |:|:  :|||||:||:||  |
Qy    309 W-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAG 367

Db    362 tfcgngpfplvhil-nellvqtesnstp 388
            |  | :||:: |  || :   :::||
Qy    368 FSCNQGRYPLIQTLRQELSLPYLPSGTP 395


RESULT    4
ID    P81342 standard; protein; 321 AA.
AC    P81342;
DT    19-OCT-1990  (first entry)
DE    Polypeptide involved in protective mechanisms
KW    Immune response; cell growth.
PN    J63032898-A.
```

```
PD   30-JAN-1988.
PF   16-JUL-1986; 167518.
PR   16-JUL-1986; JP-167518.
PA   (DAIN) Dainippon Pharm KK.
DR   WPI; 88-068419/10.
DR   P-PSDB; P81342.
PT   New polypeptide and DNA encoding it - related to protective
PT   mechanisms such as immune response etc.
PS   Disclosure; p; Japanese.
CC   This polypeptide is involved in protective mechanisms such as immune
CC   response, cell growth and activation of protective functions.
SQ   Sequence   321 AA;

  Query Match           26.5%;  Score 912;  DB 1;  Length 321;
  Best Local Similarity  49.6%;  Pred. No. 3.09e-73;
  Matches   114;  Conservative   65;  Mismatches  45;  Indels   6;  Gaps   3;

Db       7 tgfvvlvllqccsayklvcyytswsqyregdgscfpdaldrflcthiiysfanisndhid 66
Qy       8 AGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLS 67

Db      67 twewndvtlygmlntlknrnpnlktllsvggwnfgsqrfskiasntqsrrtfiksvppfl 126
Qy      68 TTEWNDETLYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFL 127

Db     127 rthgfdgvdlawlypg-rg----dkqhfttlikemkaefikeaq-pgkkqlllsaalsag 180
Qy     128 RKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG 187

Db     181 kvtidssydiakisqhldlisimtydfhgawrgttghhsplfrgqedasp 230
Qy     188 QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGA 237


RESULT    5
ID   R70025 standard; Protein; 371 AA.
AC   R70025;
DT   29-SEP-1995  (first entry)
DE   Tobacco chitinase encoded by tobacco Cluster-A cDNA clone cA-3.
KW   Tobacco; chitinase; antifungal; fungicide; Cluster-A.
OS   Nicotiana tabacum.
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT   /label= signal peptide
FT   Misc_difference 22
FT   /note= "not present in genomic clone"
FT   Misc_difference 182..195
FT   /note= "Determined by protein sequencing"
FT   Misc_difference 216..241
FT   /note= "Determined by protein sequencing"
PN   EP-639642-A.
PD   22-FEB-1995.
PF   17-AUG-1993; 202425.
PR   17-AUG-1993; EP-202425.
PA   (MOGE-) MOGEN INT NV.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PI   Apotheker-de Groot M,  Bol JF,  Cornelissen BJC,  Linthorst HJM;
PI   Melchers LS,  Ponstein AS,  Sela-buurlage MB;
DR   WPI; 95-083454/12.
DR   N-PSDB; Q82973.
PT   New plant protein having endo-chitinase activity - used in
PT   antifungal compsns. and to develop transformed plants which are
PT   less susceptible to fungal infection.
PS   Claim 3; Page 19-20; 43pp; English.
CC   Screening of a lambda ZAP cDNA library of TMV-infected Samsunn NN
CC   tobacco plants with a probe derived from PROB40 (a partial Cluster-A
CC   cDNA clone) resulted in the isolated of 11 positively hybridising
CC   clones. Analysis revealed that all were identical and corresp. to
CC   Cluster-A cDNA. The nt sequence of clone cA-3 is given in Q82973/
CC   R70025. This clone lacks 7 codons of the 5' part of the ORF. The
CC   cDNA clone was completed by performing a PCR reaction on clone cA-3
CC   with primer T7 (Q82794) and primer P1 (Q82975). A partial AA
CC   sequence of the protein isolated from TMV-infected tobacco leaves
CC   was obtd. and found to be almost identical to the AA sequence
CC   deduced from the cDNA.
SQ   Sequence   371 AA;

  Query Match           17.3%;  Score 597;  DB 13;  Length 371;
  Best Local Similarity  33.0%;  Pred. No. 2.36e-43;
  Matches   107;  Conservative   86;  Mismatches 108;  Indels  23;  Gaps  18;

Db      36 nnidstlfthlfcafadlnpqsnqliispenqds-fsqftstvqrknpsvktflsiaggr 94
Qy      44 KDLDPSLCTHLIYAFAGMT--NHQLSTTEWNDETLYQEFNG-LKKMNPKLKTLLAI-GGW 99

Db      95 a-dttaygimarqpnsrksfidssirlarqfgfhgldldweyp--l-s-atdmtnlgill 149
Qy     100 NFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLV 159

Db     150 newrtainmearnsgraallltaavsysprvnglnypvesvarnlnwinlmaydfygpnw 209
Qy     160 QDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGS-W 218

Db     210 spsqtnshaqlfdpvnh--isgs---dg-inawiqagvptkklvlgipfygyawrlvnpn 263
Qy     219 EKV-TGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSS 277

Db     264 ihdlrapaagksnvga-vddgsm-tynrirdyivqtrattvynativgdycysgsnwisy 321
Qy     278 DTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQR-IQDQK-VP-YIFRDNQWVGF 334

Db     322 ddtqsvrnkvnyvkgrgllgyfaw 345
Qy     335 DDVESFKTKVSYLKQKGLGGAMVW 358


RESULT    6
ID   R70029 standard; Protein; 377 AA.
AC   R70029;
DT   29-SEP-1995  (first entry)
DE   Tobacco cluster-A protein encoded by genomic clone.
KW   Tobacco; chitinase; antifungal; fungicide; Cluster-A.
OS   Nicotiana tabacum.
FH   Key             Location/Qualifiers
FT   Peptide         1..25
FT   /label= signal peptide
FT   Misc_difference 28..29
FT   /note= "cDNA encodes additional Ile here"
FT   Misc_difference 188..201
FT   /label= Determined by sequencing the protein
FT   Misc_difference 222..247
FT   /note= "Determined by sequencing the protein"
PN   EP-639642-A.
PD   22-FEB-1995.
PF   17-AUG-1993; 202425.
PR   17-AUG-1993; EP-202425.
PA   (MOGE-) MOGEN INT NV.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PI   Apotheker-de Groot M,  Bol JF,  Cornelissen BJC,  Linthorst HJM;
PI   Melchers LS,  Ponstein AS,  Sela-buurlage MB;
DR   WPI; 95-083454/12.
DR   N-PSDB; Q82976.
PT   New plant protein having endo-chitinase activity - used in
PT   antifungal compsns. and to develop transformed plants which are
PT   less susceptible to fungal infection.
PS   Claim 3; Page 25-28; 43pp; English.
CC   Screening of a lambda ZAP cDNA library of TMV-infected Samsunn NN
CC   tobacco plants with a probe derived from PROB40 (a partial Cluster-A
CC   cDNA clone) resulted in the isolated of 11 positively hybridising
CC   clones. Analysis revealed that all were identical and corresp. to
CC   Cluster-A cDNA.The sequence of cDNA clone cA-3 is given in Q82973/
CC   R70025. A genomic library of N. tabacum was screened using the
CC   Cluster-A cDNA insert of clone cA-3 as a probe (see Q82977, Q82978).
CC   The complete nt. sequence of Cluster-A cDNA including the deduced
CC   primary structure of the Cluster-A protein the 5' and 3' UTR regions
CC   of the gene are shown in Q82976/R70029. Comaprison of the cDNA
```

```
CC   clone with the Cluster-A gene revealed that these sequences share
CC   a high degree of identity (94%). The Cluster-A precursor protien
CC   contains a putative signal peptide as well as 4 potential N-linked
CC   glycosylation sites (N-X-S/T). The predicted mature protein has
CC   a calculated mol. wt. of 39,033 Da.
SQ   Sequence   377 AA;

  Query Match           17.3%;  Score 597;  DB 13;  Length 377;
  Best Local Similarity  33.0%;  Pred. No. 2.36e-43;
  Matches   107;  Conservative   88;  Mismatches 106;  Indels  23;  Gaps  18;

Db     42 nnidstlfthlfcafadlnpqlnqliispenqds-frqftstvqrknpsvktflsiaggr 100
          :::|::| |||: ||| :  :  ||  :  |::: :::| : : ::|| :||:|:| ||:
Qy     44 KDLDPSLCTHLIYAFAGMTNH--QLSTTEWNDETLYQEFNG-LKKMNPKLKTLLAI-GGW 99

Db    101 ansta-ygimarqpnsrksfidssirlarqlgfhgldldweyp--l-saa-dmtnlgtll 155
            :|  :  |   :|:| :|:::|:||: |  :| |||||||||    |:| |:  : ||:
Qy    100 NFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLV 159

Db    156 newrtaintearnsgraallltaavsnsprvnglnypveslarnldwinlmaydfygpnw 215
             :   |:  ||: ||:  |||:|||: :       | |: :|:||| :|||||| |: |
Qy    160 QDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGS-W 218

Db    216 spsqtnshaqlfdpvnh--vsgs---dg-inawiqagvptkklvlgipfygyawrlvnan 269
              |  :: |:   :    ::|   |: :  |:| | |: ||:||:|:| || :  | :::
Qy    219 EKV-TGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSS 277

Db    270 ihglrapaagksnvga-vddgsm-tynrirdyivesrattvynativgdycysgsnwisy 327
             : |||:| :  |:   :|:| :|  :       ::   : :   |  | :  : |:::
Qy    278 DTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQR-IQDQK-VP-YIFRDNQWVGF 334

Db    328 ddtqtvrnkvnyvkgrgllgyfaw 351
          || :: : ||:|:| :|| |   |
Qy    335 DDVESFKTKVSYLKQKGLGGAMVW 358


RESULT    7
ID   R63441 standard; Protein; 424 AA.
AC   R63441;
DT   06-JUL-1995  (first entry)
DE   Trichoderma harzianum P1 endochitinase.
KW   Trichoderma harzianum P1; endochitinase; inhibition;
KW   chitin containing fungi; herbivorous insects.
OS   Trichoderma harzianum.
PN   WO9424288-A.
PD   27-OCT-1994.
PF   10-FEB-1994; U01198.
PR   14-APR-1993; US-045269.
PA   (CORR ) CORNELL RES FOUND INC.
PI   Harman GE,  Hayes CK,  Lorito M,  Tronsmo A;
DR   WPI; 94-341867/42.
DR   N-PSDB; Q78144.
PT   Trichoderma harzianum P1 endo:chitinase and chitobiase genes -
PT   useful in biological control of chitin-contg. fungi and insects
PT   (Eng).
PS   Example 8; Page 45; 62pp; English.
CC   Q78144 encodes R63441 the Trichoderma harzianum P1 endochitinase,
CC   which can be used to inhibit chitin containing fungi and chitin
CC   containing herbivorous insects, especially from the genera below
CC   Fusarium, Gliocladium, Rhizoctonia, Trichoderma, Uncinula, Ustilage,
CC   Erysiphe, Botrytis, Saccharomyces, Sclerotium and Alternaria.
SQ   Sequence   424 AA;

  Query Match           16.9%;  Score 581;  DB 12;  Length 424;
  Best Local Similarity  36.0%;  Pred. No. 7.47e-42;
  Matches   102;  Conservative   69;  Mismatches  95;  Indels  17;  Gaps  13;

Db    112 qlfk-lkkangnlkvmlsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidv 169
          | |: ||| | :|| :|:|||| :| :|    :|  ||  | ::|| |::  :|||:|:
Qy     78 QEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 137

Db    170 dweyp-addtqatnmvlllkeirsqldayaaqy-apg-yhfllsiaapagpehy-sflhm 225
          ||||| :: : | ::  :   ::   :|:  :  ::|  ::||| | |||     :   :
Qy    138 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 197

Db    226 sdlgqvldyvnlmaydyagswssysghdanlfanpsnpnssp-yntdqaikdyikggvpa 284
            ::| ||:|||||||: |||   :||:: |:   :: :::  | | |: : :  | ||
Qy    198 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 257

Db    285 skivlgmpiygraf--estgg--igqtysgigs-gsw--engiwdy-kvlpkagatvqyd 336
          ||::|||| |||:|   |::   :|   :| |: |:   | |:  |  | :  ||| |
Qy    258 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQ-- 315

Db    337 svaqayysydpsskelisfdtpdmintkvsylknlglggsmfw 379
           : :   :|    :: ::||  : ::|||||||  ||||:| |
Qy    316 RIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT    8
ID   R56861 standard; Protein; 389 AA.
AC   R56861;
DT   17-FEB-1995  (first entry)
DE   Aphanocladium album mature endochitinase.
KW   oxalic acid degradation; oxalate oxidase; selection marker;
KW   filamentous fungus endochitinase; precursor;
KW   pathogen resistance; plant cell selection.
OS   Aphanocladium album.
FH   Key             Location/Qualifiers
FT   mat_peptide     1..1167
FT   /*tag= a
FT   /product= endochitinase
PN   WO9413790-A.
PD   23-JUN-1994.
PF   07-DEC-1993; F01203.
PR   07-DEC-1992; FR-014721.
PA   (SNFI ) ELF SANOFI.
PA   (ERAP ) SOC NAT ELF AQUITAINE.
PI   Grezes-Besset B,  Grison R,  Pignard A,  Schneider M;
DR   WPI; 94-217870/26.
DR   N-PSDB; Q68353.
PT   Use of gene encoding oxalic acid degrading enzyme - for plant
PT   cell selection, esp. gene coupled to gene-encoding protein
PT   conferring pathogen resistance
PS   Disclosure and Claim 8; Page 39-40; 64pp; French.
CC   The sequence coding for wheat germin, a protein induced during
CC   germination of wheat, is a preferred example of DNA coding for a
CC   protein able to degrade oxalic acid. Such DNA can be coupled to a
CC   sequence of interest, e.g. one coding for enzymes which confer
CC   resistance to pathogens (esp. endochitinases). Plant calli which
CC   have been successfully transformed by the sequences are selected
CC   by their ability to grow on oxalic acid-contg. medium. R56861 is
CC   a known sequence of a mature endochitinase; the sequence was
CC   previously disclosed in EP-531218.
SQ   Sequence   389 AA;

  Query Match           16.7%;  Score 576;  DB 10;  Length 389;
  Best Local Similarity  36.2%;  Pred. No. 2.20e-41;
  Matches   101;  Conservative   65;  Mismatches  95;  Indels  18;  Gaps  16;

Db     81 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 139
          ||| | ::| :|:|||| :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy     83 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 142

Db    140 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 194
           :::: | ::  :   |:   ::: :  : | ::||| | ||| : |    :  : ::
Qy    143 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 201

Db    195 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 253
          ||::|||||: ||| : |||:: :|   ::: |::  | | ||| :  | ||:|::
Qy    202 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 261

Db    254 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 305
          |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    262 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 321

Db    306 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 344
```

```
          | :    :: ::||    : |||| || |||||:| |
Qy    322 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT    9
ID   R33072 standard; Protein; 423 AA.
AC   R33072;
DT   30-JUN-1993  (first entry)
DE   Aphanocladium album pre-pro-endochitinase.
KW   filamentous fungus; pathogen resistance.
OS   Aphanocladium album.
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT   /label= signal
FT   /note= "pre-sequence"
FT   Peptide         23..34
FT   /note= "pro-sequence"
FT   Protein         35..423
FT   /label= endochitinase
PN   EP-531218-A.
PD   10-MAR-1993.
PF   04-SEP-1992; 402414.
PR   06-SEP-1991; FR-011072.
PA   (SNFI ) ELF SANOFI.
PA   (ERAP ) SOC NAT ELF AQUITAINE.
PI   Blaiseau P,  Legoux R,  Leguay J,  Schneider M;
DR   WPI; 93-078700/10.
DR   N-PSDB; Q37711, Q37712.
PT   Recombinant DNA encoding fungal endo:chitinase - useful for
PT   protecting plants against fungi, insects, etc.
PS   Claim 1; Fig 1; 75pp; French.
CC   Clone CH3C was isolated from a cDNA bank prepared from A.album
CC   mRNA by screening with anti-chitinase antiserum. Three probes were
CC   designed based on the CH3C sequence, corresponding to the start,
CC   middle and downstream regions of the coding sequence. The probes
CC   were used to isolate the full-length coding sequence of A.album
CC   endochitinase from an A.album cDNA library in pTZ19R. A 1.6kb
CC   fragment was identified in several positive clones. The sequence
CC   of the fragment contained an ORF encoding a 423 amino acid protein.
CC   The mature endochitinase can be used to control pathogens
CC   such as fungi, bacteria, arthropods and nematodes on plants.
SQ   Sequence   423 AA;

  Query Match           16.7%;  Score 576;  DB 6;  Length 423;
  Best Local Similarity  36.2%;  Pred. No. 2.20e-41;
  Matches   101;  Conservative   65;  Mismatches  95;  Indels  18;  Gaps  16;

Db    115 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 173
          ||| | ::| :|:|||| :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy     83 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 142

Db    174 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 228
           :::: | ::  :   |:   ::: :  : |  ::||| | ||| : |    :  : ::
Qy    143 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 201

Db    229 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 287
            ||::|||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    202 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 261

Db    288 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 339
          |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    262 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 321

Db    340 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 378
          | :    :: ::||    : |||| || |||||:| |
Qy    322 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT   10
ID   R33068 standard; Protein; 389 AA.
AC   R33068;
DT   30-JUN-1993  (first entry)
DE   Aphanocladium album mature endochitinase.
KW   filamentous fungus; pathogen resistance.
OS   Aphanocladium album.
PN   EP-531218-A.
PD   10-MAR-1993.
PF   04-SEP-1992; 402414.
PR   06-SEP-1991; FR-011072.
PA   (SNFI ) ELF SANOFI.
PA   (ERAP ) SOC NAT ELF AQUITAINE.
PI   Blaiseau P,  Legoux R,  Leguay J,  Schneider M;
DR   WPI; 93-078700/10.
DR   N-PSDB; Q37704.
PT   Recombinant DNA encoding fungal endo:chitinase - useful for
PT   protecting plants against fungi, insects, etc.
PS   Claim 22; Page 60-61; 75pp; French.
CC   This mature protein from the filamentous fungus Aphanocladium
CC   album has endochitinase activity and is used to control pathogens
CC   such as fungi, bacteria, arthropods and nematodes on plants.
CC   Recombinant DNA comprising the endochitinase coding sequence is
CC   specifically intended for transforming Nicotiana tabacum, Helianthus
CC   annuus and Brassica napus to pathogen resistance.
SQ   Sequence   389 AA;

  Query Match           16.5%;  Score 569;  DB 6;  Length 389;
  Best Local Similarity  35.8%;  Pred. No. 9.94e-41;
  Matches   100;  Conservative   66;  Mismatches  95;  Indels  18;  Gaps  16;

Db     81 lkkqnrnmkvmvsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 139
          ||| | ::| :::|||| :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy     83 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 142

Db    140 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 194
           :::: | ::  :   |:   ::: :  : |  ::||| | ||| : |    :  : ::
Qy    143 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 201

Db    195 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 253
            ||::|||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    202 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 261

Db    254 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 305
          |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    262 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 321

Db    306 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 344
          | :    :: ::||    : |||| || |||||:| |
Qy    322 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT   11
ID   W02159 standard; Protein; 866 AA.
AC   W02159;
DT   14-JAN-1997  (first entry)
DE   Soluble chitinase.
KW   Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;
KW   Beta-N-acetylglucosamidase; chitin; oligosaccharide; catabolic;
KW   catabolism.
OS   Vibrio furnissii.
PN   WO9625424-A1.
PD   22-AUG-1996.
PF   13-FEB-1996; U02332.
PR   13-FEB-1995; US-386727.
PA   (UYJO ) UNIV JOHNS HOPKINS.
PI   Bassler B,  Chitlaru E,  Keyhani N,  Roseman S,  Rowe C;
PI   Yu C;
DR   WPI; 96-393335/39.
DR   N-PSDB; T36390.
PT   Chitin biosynthetic enzymes end I, exo I and exo II  - are
PT   periplasmic chito:dextrinase(s), periplasmic beta-GlcNAcidase(s) and
PT   aryl beta-N-acetyl:gluco:amidase(s), respectively
PS   Example 4; Page 79-82; 101pp; English.
CC   Periplasmic chitodextrinase (W02156), periplasmic
CC   Beta-N-acetylglucosaminidase (W02157) and aryl
CC   Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin
CC   oligosaccharides with the structure (GlcNAc)n where n is 2 or
```

```
CC   higher, by contacting them with soluble chitin.  The enzymes are
CC   encoded by the genes endI, exoI and exoII respectively.  They are
CC   all genes involved in the catabolic pathway of chitin.
SQ   Sequence   866 AA;

  Query Match           13.4%;  Score 463;  DB 19;  Length 866;
  Best Local Similarity  30.7%;  Pred. No. 7.19e-31;
  Matches    93;  Conservative   81;  Mismatches 104;  Indels  25;  Gaps  20;

Db    253 fgnlmalkkanpnlkilpsvggwtlsdp-fy-ffsdktkrdtfvasmkeylqtwkffdgv 310
          : :: :||| ||:|| | ::||| ::   |    :   :|:||| |   :|:   | ||:
Qy     77 YQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSF-DGL 135

Db    311 didwefpggqganpnlggpndgatyvalmkelramldeleaet-grqyeltsaisaggdk 369
          |:|||:||:||: | :   ::   : :|: :|   ::: ||:| |::  | ||  ::|:
Qy    136 DLDWEYPGSQGS-PAVD--KE--RFTTLVQDLANAFQQ-EAQTSGKERLLLSAAVPAGQT 189

Db    370 iakvdyqaaq--qymdyiflmsydfsgafdlknlahqtnly--asswdpatkyttdkgvk 425
                |:     | :|:: ||:||| |: : |  :| : ||       :|:    | :|
Qy    190 YVDAGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNSPLYKRQEESGAAASLNVDAAVQ 248

Db    426 allgqgvtpgkvvvgaamygrgwt-gvng-yqagnpftgsat-gpikgtwengvvdyrdi 482
            |  |  ::|:::| : |||: | : ::  : | | |||:| ||:    | |:: | ::
Qy    249 QWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFT-K-EGGMLAYYEV 306

Db    483 vnnrmgagweqgydetaeapyvfkastgdlisfdndrsvkakgqyvlanqlgglfaweid 542
          : ::||   |  ::    ||:|: :  : ::||:  | |:|  |:  : |||   | :|
Qy    307 CS-WKGAT-KQRIQD-QKVPYIFRDN--QWVGFDDVESFKTKVSYLKQKGLGGAMVWALD 361

Db    543 adn 545
            |:
Qy    362 LDD 364


RESULT   12
ID   R64823 standard; Protein; 561 AA.
AC   R64823;
DT   23-JUL-1995  (first entry)
DE   Chitinase A.
KW   Chitinase A; protein secretion; biological control agent;
KW   transgenic plant; chitin; plant pathogen; nematode; fungus;
KW   crop protection; pCHIT1251.
OS   Serratia marcescens QMB1466.
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT   /label= Sig_peptide
PN   US5374540-A.
PD   20-DEC-1994.
PF   26-MAR-1984; 593691.
PR   26-MAR-1984; US-593691.
PR   18-JUL-1986; US-888033.
PR   09-JUL-1990; US-550253.
PA   (DNAP ) DNA PLANT TECHNOLOGY CORP.
PI   Jones JDG,  Suslow T;
DR   WPI; 95-035648/05.
DR   N-PSDB; Q76290.
PT   Plant cell contg. signal sequence of chitinase A gene - for
PT   secretion of foreign polypeptide, esp. chitinase A for protection
PT   against pathogenic fungi and nematodes.
PS   Disclosure; Fig. 1A-1C; 20pp; English.
CC   Serratia marcescens QMB1466 genomic DNA was partially digested,
CC   ligated into vector pLAFR1, packaged into lambda phage particles,
CC   and used to transfect Escherichia coli DH1 (ATCC 33849) cells.
CC   Plating on chitin medium was used to screen for chitinase activity.
CC   Clone C12 (ATCC 67152) was isolated.  Its sequence is given in
CC   Q76290, and the deduced protein sequence in R64823.
SQ   Sequence   561 AA;

  Query Match           11.3%;  Score 389;  DB 12;  Length 561;
  Best Local Similarity  28.2%;  Pred. No. 4.43e-24;
  Matches    91;  Conservative   81;  Mismatches 126;  Indels  25;  Gaps  18;

Db    231 pfaalqkaqkgvtawddpykgnfgqlmalkqahpdlkilpsiggwtlsdpfff-mgdkvk 289
          :||:: : | : | |:|     : :: :||  :| || | :|||| ::   |  |    :
Qy     57 AFAGMTNHQLSTTEWNDE-T-LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATAN 114

Db    290 -rdrfvgsvkeflqtwkffdgvdidwefpggkganpnlgspqdgetyvllmkelramldq 348
           |: || |   ||:   | ||:|:|||:||: |: | :    | | :  |: :|   ::|
Qy    115 NRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV----DKERFTTLVQDLANAFQQ 168

Db    349 lsaetgr-kyeltsaisagkdkidkvaynvaq--nsmdhiflmsydfygafdlknlghqt 405
          :  :|: :  |::|::||   :|  :|:|     ::| : ||:||| |: : |  || :
Qy    169 EAQTSGKERLLLSAAVPAGQTYVD-AGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNS 226

Db    406 alnar---pgsrhrlhhgerrecaagqgvkpgkivvgtamygrgwtgvngyqnnipftgt 462
          :|  |   :|:   |:     :     |  ::|:::| : |||: |  :: :  :   :|
Qy    227 PLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPAT 286

Db    463 hravkgtw--engivdyrqiasqfmsgewqytydataeapyvfkpstgdlitsddarsvq 520
            :  |    | |:: | :: |  :::  |   :     ||:|:    : :  ||  |
Qy    287 GSGTPGPFTKEGGMLAYYEVCS-WKGATKQ-RIQD-QKVPYIFR-DN-QWVGFDDVESFK 341

Db    521 akgkyvldkqlgglfsweidadn 543
          :|  |: :| |||   | :| |:
Qy    342 TKVSYLKQKGLGGAMVWALDLDD 364


RESULT   13
ID   R48981 standard; Protein; 561 AA.
AC   R48981;
DT   04-SEP-1994  (first entry)
DE   Bacterial chitinase-A fragment from pCHIT1251.
KW   Chitinase-A; enzyme; transgenic plant; biological control;
KW   disease-resistance; crop improvement; Agrobacterium; fungus;
KW   nematode; insect.
PN   US5290687-A.
PD   01-MAR-1994.
PF   26-MAR-1984; 593691.
PR   26-MAR-1984; US-593691.
PR   18-JUL-1986; US-888033.
PR   09-JUL-1990; US-550253.
PR   14-AUG-1992; US-930970.
PA   (DNAP ) DNA PLANT TECHNOLOGY CORP.
PI   Jones JDG,  Suslow T;
DR   WPI; 94-074348/09.
DR   N-PSDB; Q57966.
PT   Inhibition of chitinous fungal plant pathogens - by transforming
PT   plant with DNA encoding chitinase activity
PS   Disclosure; Fig.1; 13pp; English.
CC   Plasmid pCHIT1251 containing the chitinase-A gene (chiA), fused
CC   directly to a nos promoter,  may be used to transform, via
CC   Agrobacterium sp., a crop plant.  The resulting transgenic plants
CC   expressing the chitinase-A protein fragment are protected against
CC   fungi, nematodes, insects and disease agents. The protein is
CC   preferably expressed by the plant only at the site of pathogen
CC   attack, e.g. only in the root cells.  The plants express the
CC   chitinase in a biologically active form.
SQ   Sequence   561 AA;

  Query Match           10.8%;  Score 372;  DB 9;  Length 561;
  Best Local Similarity  27.6%;  Pred. No. 1.56e-22;
  Matches    89;  Conservative   84;  Mismatches 125;  Indels  25;  Gaps  18;

Db    231 pfaalqkaqkgvtawddpykgnfgqlmalkqahpdlkilpsiggwtlsdpfff-mgdkvk 289
          :||:: : | : | |:|     : :: :||  :| || | :|||| ::   |  |    :
Qy     57 AFAGMTNHQLSTTEWNDE-T-LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATAN 114

Db    290 -rdrfvgsvkeflqswkffdgvdidwefpggkganpnlgspqdgetyvlvmkevgamldq 348
           |: || |   ||:   | ||:|:|||:||: |: | :    | | :  :: :::  ::|
Qy    115 NRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV----DKERFTTLVQDLANAFQQ 168

Db    349 lwaetgr-kyeltsaisagkdkidkvaynvaq--nsmdhiflmsydfygafdlknlghqt 405
              :|: :  |::|::||   :|  :|:|     ::| : ||:||| |: : |  || :
Qy    169 EAQTSGKERLLLSAAVPAGQTYVD-AGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNS 226

Db    406 alnar---pgsrhrlhhgegrecaagqgvkpgkivvgtamygrgwtgvngyqnnipftgt 462
```

```
          :|  |   :|:   |:   : :     |  ::|:::| : |||: |  :: :  :   :|
Qy    227 PLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPAT 286

Db    463 hravkgtw--engivdyrqiasqfmsgewqytydgtaeapyvfkpstgdlitsddarsvq 520
          :  |    | |:: | :: |  :::  |   :     ||:|:    : :  ||  |
Qy    287 GSGTPGPFTKEGGMLAYYEVCS-WKGATKQ-RIQD-QKVPYIFR-DN-QWVGFDDVESFK 341

Db    521 akgkyvldkqlgglfsweidadn 543
          :|  |: :| |||   | :| |:
Qy    342 TKVSYLKQKGLGGAMVWALDLDD 364


RESULT   14
ID   R32547 standard; Protein; 238 AA.
AC   R32547;
DT   17-JUN-1993  (first entry)
DE   HaNPV IE-1 gene protein (partial sequence).
KW   Heliothis armigera nuclear polyhedrosis virus; Ha; NPV;
KW   insecticide; immediate early-1.
OS   Heliothis armigera nuclear polyhedrosis virus.
PN   WO9303144-A.
PD   18-FEB-1993.
PF   05-AUG-1992; AU0413.
PR   05-AUG-1991; AU-007576.
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PI   Christian PD, Dall DJ, Gordon KHJ, Hanzlik TN, Sriskantha A.
DR   WPI; 93-076501/09.
DR   N-PSDB; Q35982.
PT   Insect virus with reduced capacity to occlude viral particles - used
PT   for controlling proliferation of insect pests without horizontal
PT   transmission
PS   Disclosure; Fig 2; 51pp; English.
CC   The sequence is that encoded by the Heliothis armigera (Ha)
CC   nuclear polyhedrosis virus (NPV) immediate early-1 (IE-1) gene
CC   (partial sequence).
SQ   Sequence   238 AA;

  Query Match             7.9%;  Score 271;  DB 6;  Length 238;
  Best Local Similarity  34.1%;  Pred. No. 1.65e-13;
  Matches    58;  Conservative   43;  Mismatches  51;  Indels  18;  Gaps  14;

Db     44 erdkflsnpgwrefvltwkf-fdgididwefpggkganpnvgdverdnntyiallgelra 102
          :|: |::::  | | | :|: |||:|:|||:||: |: | | | ||    : :|: :|
Qy    115 NRQTFVNSAI-R-F-L-RKYSFDGLDLDWEYPGSQGS-PAV-DKER----FTTLVQDLAN 164

Db    103 mldq-vqiqtnrtlelttaisagidkiaav-ncdraqqyldkifvmsydfkgawsntdlg 160
           ::|  |   :  | |::|::||   : |  : |:  | || : :|:|||  |:| :   |
Qy    165 AFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVT-G 223

Db    161 hqtaly--gsawkpnepytanvavdallaqrvnp-kklvlgvamygrgwt 207
          | ::||       :  :   : ||:  | |:  |  ||:||:: |||: |
Qy    224 HNSPLYKRQEESGAAASLNVDAAVQQWL-QKGTPASKLILGMPTYGRSFT 272


RESULT   15
ID   W02156 standard; Protein; 1046 AA.
AC   W02156;
DT   14-JAN-1997  (first entry)
DE   Periplasmic chitodextrinase.
KW   Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;
KW   Beta-N-acetylglucosamidase; chitin; oligosaccharide; catabolic;
KW   catabolism.
OS   Vibrio furnissii.
PN   WO9625424-A1.
PD   22-AUG-1996.
PF   13-FEB-1996; U02332.
PR   13-FEB-1995; US-386727.
PA   (UYJO ) UNIV JOHNS HOPKINS.
PI   Bassler B,  Chitlaru E,  Keyhani N,  Roseman S,  Rowe C;
PI   Yu C;
DR   WPI; 96-393335/39.
DR   N-PSDB; T36387.
PT   Chitin biosynthetic enzymes end I, exo I and exo II  - are
PT   periplasmic chito:dextrinase(s), periplasmic beta-GlcNAcidase(s) and
PT   aryl beta-N-acetyl:gluco:amidase(s), respectively
PS   Claim 2; Page 68-71; 101pp; English.
CC   Periplasmic chitodextrinase (W02156), periplasmic
CC   Beta-N-acetylglucosaminidase (W02157) and aryl
CC   Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin
CC   oligosaccharides with the structure (GlcNAc)n where n is 2 or
CC   higher, by contacting them with soluble chitin.  The enzymes are
CC   encoded by the genes endI, exoI and exoII respectively.  They are
CC   all genes involved in the catabolic pathway of chitin.
SQ   Sequence   1046 AA;

  Query Match             6.4%;  Score 219;  DB 19;  Length 1046;
  Best Local Similarity  33.3%;  Pred. No. 5.16e-09;
  Matches    48;  Conservative   40;  Mismatches  42;  Indels  14;  Gaps  10;

Db    454 etfadsavemmrkyrfdgldidlrisnidggtgnpddtafsesrraylmnsyhelmrvlr 513
          :|| :||:  :||| ||||||:| :  :  |: |:|      | | : |:   ::|   ::
Qy    117 QTFVNSAIRFLRKYSFDGLDLD-W--EYPGSQGSPA-VD-KE-RFTTLV---QDLANAFQ 167

Db    514 ekldvasaqdgvhymltiaapss-ayllrgmetmavtqyldyvnimsydlhgawndhvgh 572
          :  : :|: :  : :|: | |:: :|:  | |   ::| ||:||:|:||:||:|:   ||
Qy    168 QEAQ-TSGKE--RLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGH 224

Db    573 naalydtgkdselaqw-nvygtaq 595
          |::||    :|  |   || :: |
Qy    225 NSPLYKRQEESGAAASLNVDAAVQ 248


Search completed: Fri Jun 27 16:54:04 1997
Job time : 116 secs.
```

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 27 16:49:12 1997; MasPar time 21.48 Seconds
618.601 Million cell updates/sec
Tabular output not generated.

Title: >US-08-663-618A-2
Description: (1-466) from US08663618A.pep
Perfect Score: 3443
Sequence: 1 MVRSVAWAGFMVLLMIPWGS.....QQSCPTGLVFSNSCKCCTWN 466

Scoring table: PAM 150
Gap 11

Searched: 89912 seqs, 28507787 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir50
1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8
13:unann9 14:unann10 15:unenc 16:unrev

Statistics: Mean 48.408; Variance 102.872; scale 0.471

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1519	44.1	383	13	A49562	cartilage glycoprote	8.84e-265
2	1462	42.5	383	16	S51327	heparin-binding glyc	1.71e-253
3	1459	42.4	381	14	S61550	BRP39 protein - mous	6.73e-253
4	1459	42.4	381	14	I48271	BRP39 protein - mous	6.73e-253
5	1380	40.1	399	5	S27879	secretory protein YM	2.80e-237
6	1361	39.5	539	14	I46470	estrogen dependent o	1.58e-233
7	1345	39.1	537	14	S57197	oviduct-specific gly	2.27e-230
8	1185	34.4	483	12	A53918	chitinase (EC 3.2.1.	7.24e-199
9	1086	31.5	504	12	A38221	chitinase (EC 3.2.1.	1.85e-179
10	1031	29.9	554	12	A56596	chitinase (EC 3.2.1.	1.04e-168
11	698	20.3	699	10	A38368	chitinase (EC 3.2.1.	2.46e-104
12	594	17.3	378	5	S51591	chitinase (EC 3.2.1.	1.33e-84
13	582	16.9	424	5	S47133	chitinase (EC 3.2.1.	2.42e-82
14	576	16.7	423	5	JQ1975	chitinase (EC 3.2.1.	3.26e-81
15	561	16.3	423	12	S51369	chitinase - fungus (	2.14e-78
16	546	15.9	427	12	JC4565	chitinase (EC 3.2.1.	1.38e-75
17	446	13.0	563	16	S60651	chitinase precursor	4.70e-57
18	404	11.7	561	9	A25090	chitinase (EC 3.2.1.	2.11e-49
19	344	10.0	499	9	S52422	chitinase (EC 3.2.1.	1.19e-38
20	345	10.0	799	8	PC4106	chitinase (EC 3.2.1.	7.93e-39
21	344	10.0	820	16	A40633	chitinase (EC 3.2.1.	1.19e-38
22	337	9.8	499	9	S04856	chitinase (EC 3.2.1.	2.06e-37
23	314	9.1	511	11	S61166	probable membrane pr	2.24e-33
24	287	8.3	610	10	JH0573	chitinase (EC 3.2.1.	1.06e-28
25	266	7.7	452	12	JC4038	47K glycoprotein - f	4.02e-25
26	263	7.6	1146	11	S07915	RF2 protein - yeast	1.29e-24
27	246	7.1	413	5	JC2135	chitinase (EC 3.2.1.	9.11e-22
28	229	6.7	124	5	S57715	chitinase (EC 3.2.1.	5.81e-19
29	172	5.0	36	14	A27682	39K whey protein - b	5.52e-10
30	172	5.0	83	13	A37954	estradiol-stimulated	5.52e-10
31	136	4.0	29	14	S57204	oviduct-specific gly	8.32e-05
32	136	4.0	29	14	S55693	oviduct-specific sia	8.32e-05
33	121	3.5	597	5	S33848	chitinase (EC 3.2.1.	8.07e-03
34	121	3.5	597	5	S32039	chitinase (EC 3.2.1.	8.07e-03
35	112	3.3	642	16	S37623	homologies with orf	1.08e-01
36	110	3.2	488	10	A41961	chitinase (EC 3.2.1.	1.90e-01
37	104	3.0	305	9	S28440	probable coproporphy	9.80e-01
38	104	3.0	422	4	ZWECC	umuC protein - Esche	9.80e-01
39	102	3.0	662	11	S61193	probable membrane pr	1.67e+00
40	105	3.0	787	8	PC1232	retrovirus-related c	7.49e-01
41	105	3.0	1409	4	OFFFCP	retrovirus-related p	7.49e-01
42	101	2.9	562	3	KRHUEB	keratin, 56K type II	2.17e+00
43	101	2.9	958	1	JN0102	glucan 1,4-alpha-glu	2.17e+00
44	101	2.9	1045	2	SUSEMM	serine proteinase (E	2.17e+00
45	101	2.9	1070	11	S19686	alpha-glucosidase (E	2.17e+00

ALIGNMENTS

RESULT 1
ENTRY A49562 #type complete
TITLE cartilage glycoprotein gp39 precursor - human
ALTERNATE_NAMES 39K synovial protein
ORGANISM #formal_name Homo sapiens #common_name man
DATE 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 03-May-1996
ACCESSIONS A49562; S10677; A33162
REFERENCE A49562
#authors Hakala, B.E.; White, C.; Recklies, A.D.
#journal J. Biol. Chem. (1993) 268:25803-25810
#title Human cartilage gp-39, a major secretory product of articular chondrocytes and synovial cells, is a mammalian member of a chitinase protein family.
#accession A49562
##status preliminary
##molecule_type mRNA
##residues 1-383 ##label HAK
##cross-references GB:M80927
REFERENCE S10677
#authors Nyirkos, P.; Golds, E.E.
#journal Biochem. J. (1990) 269:265-268
#title Human synovial cells secrete a 39 kDa protein similar to a bovine mammary protein expressed during the non-lactating period.
#cross-references MUID:90328983
#accession S10677
##molecule_type protein
##residues 22-40,'X',42-45 ##label NY2
KEYWORDS cartilage; extracellular protein; glycoprotein
FEATURE
1-21 #domain signal sequence #status predicted #label SIG\
22-383 #product cartilage glycoprotein gp39 #status predicted #label MAT
SUMMARY #length 383 #molecular-weight 42613 #checksum 2942

Query Match 44.1%; Score 1519; DB 13; Length 383;
Best Local Similarity 52.8%; Pred. No. 8.84e-265;
Matches 201; Conservative 88; Mismatches 83; Indels 9; Gaps 5;

```
Db      8 tgfvvlvllqccsayklvcyytswsqyregdgscfpdaldrflcthiiysfanisndhid 67
          :||:||:::   || |||||:|:|:|||:|::  :|  ||  ||||:||:|| ::| ::
Qy      8 AGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLS 67

Db     68 twewndvtlygmlntlknrnpnlktllsvggwnfgsqrfskiasntqsrrtfiksvppfl 127
```

```
          | ||||  |||  :| ||: ||:||||||::||||||:|:|: : : : :|:||::|   ||
Qy     68 TTEWNDETLYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFL 127

Db    128 rthgfdgldlawlypgrr-----dkqhfttlikemkaefikeaq-pgkkqlllsaalsag 181
          |  :|||||| | ||| :     ||::|||||: ::    |  ||| :|| :|||||||::||
Qy    128 RKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG 187

Db    182 kvtidssydiakisqhldfisimtydfhgawrgttghhsplfrgqedaspdrfsntdyav 241
             :|::|:: ||:|:|||:::|:|||||:|   |||:|||:: ||::::    | | ||
Qy    188 QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAV 247

Db    242 gymlrlgapasklvmgiptfgrsftlass-etgvgapisgpgipgrftkeagtlayyeic 300
             |: |:|||||::|:||:|||||||||| :| |||| :|:| || ||||:| |||||:|
Qy    248 QQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVC 307

Db    301 dflrgatvhrtlgqqvpyatkgnqwvgyddqesvkskvqylkdrqlagamvwaldlddfq 360
             :||| :|   | |||  : |||||:|| || |:|| |||:: |:|||||||||||
Qy    308 SW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFA 366

Db    361 gsfcgqdlrfpltnaikdala 381
          |  | |  |:||  :::: |:
Qy    367 GFSCNQG-RYPLIQTLRQELS 386


RESULT    2
ENTRY             S51327     #type complete
TITLE             heparin-binding glycoprotein 38K - pig
ORGANISM          #formal_name Sus scrofa domestica #common_name domestic pig
DATE              07-May-1995 #sequence_revision 07-May-1995 #text_change
                    07-May-1995
ACCESSIONS        S51327
REFERENCE         S51327
   #authors       Shackelton, L.M.; Mann, D.M.; Millis, A.J.T.
   #submission    submitted to the EMBL Data Library, January 1995
   #description   Identification of a 38kDa heparin-binding glycoprotein
                    (gp38k) in differentiating vascular smooth muscle cells as
                    a member of a group of proteins associated with tissue
                    remodeling.
   #accession     S51327
      ##status         preliminary
      ##residues       1-383 ##label SHA
      ##cross-references EMBL:Z47803
SUMMARY           #length 383  #molecular-weight 42443  #checksum 2907

  Query Match           42.5%;  Score 1462;  DB 16;  Length 383;
  Best Local Similarity  51.0%;  Pred. No. 1.71e-253;
  Matches   198;  Conservative   87;  Mismatches  94;  Indels   9;  Gaps   5;

Db      1 mglrvaqtgfvalvllqscaayklvcyytswsqyregdgscfpdaidpflcthiiysfan 60
            |   || :||: |:::   :| |||||:|:|:|||:|::  :|  :|| ||||:||:||
Qy      1 MVRSVAWAGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAG 60

Db     61 isnneidtlewndvtlydtlntlknrnpnlktllsvggwnfgsqrfskiasntqsrrtfi 120
          ::|::: | |||| |||: :| ||: ||:|||||::|||||:|:|: : : : :|:|||:
Qy     61 MTNHQLSTTEWNDETLYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFV 120

Db    121 ksvppflrthgfdgldlawispgrr-----dkrhlttlvkemkaefvreal-pgterlll 174
          :|   |||  :|||||| |  || :     || ::|||| ::   | :||  :| |||||
Qy    121 NSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLL 180

Db    175 sgavsagkvaidrgydiaqisqhldfislltydfhgawrqttghhsplfrgqgdassdrf 234
          |:||:||   :| ||::  |:|:|||::|::|||||:|   |||:|||:: | ::::
Qy    181 SAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAAS 240

Db    235 snadyavsyvlrlgapanklvmgiptfgrsftlassk-tdvgapasgpgipgrftkekgi 293
            | | ||   |: |:||:||::|:||:|||||||||  | |||||:|:| || |||| |:
Qy    241 LNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGM 300

Db    294 layyeicdflqgatvrrplgqqvpyatkgnqwvgyddqesvknkakylksrqlagamvwt 353
          |||||:|    ||| :|   | |||  : |||||:|| || | |  ||| : |:||||:
Qy    301 LAYYEVCSW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWA 359

Db    354 ldlddfrgnfcgqnlrfpltsaikdvla 381
          |||||| |  | |  |:||  :::: |:
Qy    360 LDLDDFAGFSCNQG-RYPLIQTLRQELS 386


RESULT    3
ENTRY             S61550     #type complete
TITLE             BRP39 protein - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change
                    03-May-1996
ACCESSIONS        S61550
REFERENCE         S61550
   #authors       Morrison, B.W.
   #submission    submitted to the EMBL Data Library, November 1995
   #accession     S61550
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-381 ##label MOR
      ##cross-references EMBL:X93035
SUMMARY           #length 381  #molecular-weight 43001  #checksum 5107

  Query Match           42.4%;  Score 1459;  DB 14;  Length 381;
  Best Local Similarity  49.9%;  Pred. No. 6.73e-253;
  Matches   194;  Conservative   88;  Mismatches  96;  Indels  11;  Gaps   9;

Db      1 mgmraaltgfavlmlllqscsayklvcyftswsqyregvgsflpdaiqpflcthiiysfan 60
          |    | :|| ||:::   || |||||||:|:|||:| : |||  ::| ||||:||:||
Qy      1 MVRSVAWAGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAG 60

Db     61 issdnmlstwewndesnydklnklktrntnlktllsvggwkfgekrfseiasnterrtaf 120
          :: :: ||| ||||||: |: :| ||  | :||||||:|||:||  :|::: : :: | :|
Qy     61 MT-NHQLSTTEWNDETLYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTF 119

Db    121 vrsvapflrsygfdgldlawlyp--r-lr--dkqyfstlikelnaeftkevq-pgrekll 174
          | |   ||| |:||||||| | ||  :     ||: |:||: :|   |  | | :|:|:||
Qy    120 VNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLL 179

Db    175 lsaalsagkvaidtgydiaqiaqhldfinlmtydfhgvwrqitghhsplfqgqkdtrfdr 234
          ||||::||   :|:||::  |||:|||:|||:||||| |  :|||:|||:  | ::
Qy    180 LSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA 239

Db    235 ysnvnyavqymirlgaqaskllmgiptfgksftlass-enqlgapisgeglpgrftkeag 293
            ||: |||  :: |: |||||:|:||:|:|||||| : ::||| :| | || ||||:|
Qy    240 SLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGG 299

Db    294 tlayyeicdflkgaevhrlsnekvpfatkgnqwvgyehkesvknkvgflkekklagamvw 353
          |||||:|   |||  :|: ::|||:  : |||||::  || | ||::|||:| |:|||||
Qy    300 MLAYYEVCSW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358

Db    354 aldlddfqg-tcqpkeffpltnaikdala 381
          ||||||| | :|    : ||  :::: |:
Qy    359 ALDLDDFAGFSCNQGRY-PLIQTLRQELS 386


RESULT    4
ENTRY             I48271     #type complete
TITLE             BRP39 protein - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
                    02-Jul-1996
ACCESSIONS        I48271
REFERENCE         I48271
   #authors       Morrison, B.W.; Leder, P.
   #journal       Oncogene (1994) 9:3417-3426
   #title         neu and ras initiate murine mammary tumors that share genetic
                    markers generally absent in c-myc and int-2-initiated
                    tumors.
   #cross-references MUID:95060797
   #accession     I48271
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-381 ##label RES
      ##cross-references EMBL:X93035; NID:g1085065; CDS_PID:g1085066
```

```
GENETICS
   #note          gene name brp39
SUMMARY           #length 381  #molecular-weight 43001  #checksum 5107

  Query Match           42.4%;  Score 1459;  DB 14;  Length 381;
  Best Local Similarity  49.9%;  Pred. No. 6.73e-253;
  Matches   194;  Conservative   88;  Mismatches  96;  Indels  11;  Gaps   9;

Db       1 mgmraaltgfavlmllqscsayklvcyftswsqyregvgsflpdaiqpflcthiiysfan 60
           |    | :|| ||:::   || |||||||:|:|||:| : |||  ::| ||||:||:|| 
Qy       1 MVRSVAWAGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAG 60

Db      61 issdnmlstwewndesnydklnklktrntnlktllsvggwkfgekrfseiasnterrtaf 120
           :: :: ||| ||||| |: :| ||  | :|||||::|||:||  :|::: : :: | :| 
Qy      61 MT-NHQLSTTEWNDETLYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTF 119

Db     121 vrsvapflrsygfdgldlawlyp--r-lr--dkqyfstlikelnaeftkevq-pgrekll 174
           | |   ||| |:|||||| | ||  :     ||: |:||: :|   |  | | :|:|:|| 
Qy     120 VNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLL 179

Db     175 lsaalsagkvaidtgydiaqiaqhldfinlmtydfhgvwrqitghhsplfqgqkdtrfdr 234
           ||||::||   :|:||::  |||:|||:|||:||||| |  :|||:|||:  | ::
Qy     180 LSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA 239

Db     235 ysnvnyavqymirlgaqasklllmgiptfgksftlass-enqlgapisgeglpgrftkeag 293
             ||: |||  :: |: |||||::|:||:|:|||||| : ::||| :| | || ||||:| 
Qy     240 SLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGG 299

Db     294 tlayyeicdflkgaevhrlsnekvpfatkgnqwvgyehkesvknkvgflkekklagamvw 353
           |||||:|   |||  :|: ::|||:  : |||||::  || | ||::||:| |:|||||
Qy     300 MLAYYEVCSW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358

Db     354 aldlddfqg-tcqpkeffpltnaikdala 381
           ||||||| | :|    : ||  :::: |:
Qy     359 ALDLDDFAGFSCNQGRY-PLIQTLRQELS 386


RESULT    5
ENTRY             S27879     #type complete
TITLE             secretory protein YM-1 precursor - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
                    03-May-1996
ACCESSIONS        S27879
REFERENCE         S27879
   #authors       Chang, N.C.A.; Liu, C.H.; Chang, A.C.
   #submission    submitted to the EMBL Data Library, June 1992
   #description   Molecular characterization of a secretory protein (YM-1)
                    transiently expressed by activated murine peritoneal
                    macrophages.
   #accession     S27879
      ##molecule_type mRNA
      ##residues      1-399 ##label CHA
      ##cross-references EMBL:M94584
CLASSIFICATION    #superfamily Streptomyces chitinase chi40
FEATURE
   1-21               #domain signal sequence #status predicted #label SIG\
   22-399             #product secretory protein YM-1 #status predicted #label
                        MAT
SUMMARY           #length 399  #molecular-weight 44528  #checksum 3208

  Query Match           40.1%;  Score 1380;  DB 5;  Length 399;
  Best Local Similarity  46.4%;  Pred. No. 2.80e-237;
  Matches   181;  Conservative   94;  Mismatches 112;  Indels   3;  Gaps   3;

Db       1 maklilvtglaillnvqlgssyqlmcyytswakdrpiegsfkpgnidpclcthliyafag 60
           | : :  :|: :|| :  ||:  |:||:|:||  |  |: | | ::|| ||||||||||
Qy       1 MVRSVAWAGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAG 60

Db      61 mqnneitytheqdlrdyealnglkdkkntelktllaiggwkfgpapfsamvstpqnrqif 120
           | |:::: |   |   |: :|||| |:|  ||||||||||:||   |: ||:|: ||| |
Qy      61 MTNHQLSTTEWNDETLYQEFNGLK-KMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTF 119

Db     121 iqsvirflrqynfdglnldwqypgsrgsppkdkhlfsvlvkemrkafeeesvekdiprll 180
           : | ||||| |:||||:|||:||||:|||: ||  |: || :: :||::|:      |||
Qy     120 VNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLL 179

Db     181 ltstgagiidviksgtrslnclslldyiqvmtydlhdpkdgytgensplykspydigksa 240
           |::: ::    : :|    :   ||:: :|:||:| : :  || ||||||   : | :|
Qy     180 LSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA 239

Db     241 dlnvdsiisywkdhgaaseklivgfpayghtfilsdpsktgigaptistgppgkytdesg 300
           ||||: :  | : |::: ||||:| |:||::| |: :| | :|||: ::| || :| |:|
Qy     240 SLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGG 299

Db     301 llayyevctflnegatevwdapqevpyayqgnewvgydnvrsfklkaqwlkdnnlggavv 360
           :|||||||:  : |||  :   | ||| :: |:|||:|:| ||| |   ||:: ||||:|
Qy     300 MLAYYEVCSW-K-GATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMV 357

Db     361 wpldmddfsgsfchqrhfpltstlkgdlni 390
           |:|||:|||:|  |:| ::||  ||: :|::
Qy     358 WALDLDDFAGFSCNQGRYPLIQTLRQELSL 387


RESULT    6
ENTRY             I46470     #type complete
TITLE             estrogen dependent oviduct protein precursor - sheep
ORGANISM          #formal_name Ovis orientalis aries, Ovis ammon aries
                    #common_name domestic sheep
DATE              16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change
                    16-Aug-1996
ACCESSIONS        I46470
REFERENCE         I46470
   #authors       DeSouza, M.M.; Murray, M.K.
   #journal       Endocrinology (1995) 136:2485-2496
   #title         An estrogen-dependent secretory protein, which shares
                    identity with chitinases, is expressed in a temporally and
                    regionally specific manner in the sheep oviduct at the time
                    of fertilization and embryo development.
   #cross-references MUID:95269691
   #accession     I46470
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-539 ##label DES
      ##cross-references EMBL:U16719; NID:g885600; CDS_PID:g885601
SUMMARY           #length 539  #molecular-weight 59535  #checksum 2120

  Query Match           39.5%;  Score 1361;  DB 14;  Length 539;
  Best Local Similarity  48.8%;  Pred. No. 1.58e-233;
  Matches   189;  Conservative   85;  Mismatches 106;  Indels   7;  Gaps   5;

Db       1 mgkllwvgllmlkhhdgaahklvcyftnwafsrpgsasilprdldpflcthlvfafas 60
           | : : | |::::|:   |:| ||||||||||  | | | :||:|||| |||||::|||:
Qy       1 MVRSVAWAGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAG 60

Db      61 mnnnqivpkdpldekilypefnklkernrglktllsvggwnfgtsrftkmlstfsnrerf 120
           | |:|:   :  ::  || ||| ||  |  |||||::|||||||| :|| |::| :||: |
Qy      61 MTNHQLSTTE-WNDETLYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTF 119

Db     121 vksviallrthgfdgldlfflypglrgspardrwtfvflleellqafkneaqltmrprll 180
           |:| | :||  :||||||   ||| :|||| |:  |  |:::|  ||  ||| : : |||
Qy     120 VNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLL 179

Db     181 lsaavsgdphviqkaydarllgrlldfisvlsydlhgswekvtghnsplfslpgdpkssa 240
           ||||||::   :: :|:   ::: |||:::::||:|||||||||||||:   :: ::|
Qy     180 LSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA 239

Db     241 -yam--sy--wrqlgvppekllmglptygrtfhllrasqnelgagaagpaspgkytkqag 295
            :  :   | | | |: ||::|:||||||:| |  :|:  :|| |:|:::|| :|||::|
Qy     240 SLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGG 299

Db     296 flayyevcsfvqrakkrwindqyvpyafkgkewvgyddaisfgykaffikrehfggamvw 355
           ||||||||    | |::| || ||| |: ::|||:||  ||  |  ::|:  :||||||
Qy     300 MLAYYEVCSW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358

Db     356 tldlddfrgnfcgtgpfplahtlnnll 382
```

```
          :|||||| |  |  | :|| :||   |
Qy    359 ALDLDDFAGFSCNQGRYPLIQTLRQEL 385


RESULT    7
ENTRY             S57197     #type fragment
TITLE             oviduct-specific glycoprotein 95K precursor - bovine
                    (fragment)
ORGANISM          #formal_name Bos primigenius taurus #common_name cattle
DATE              28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
                    03-May-1996
ACCESSIONS        S57197
REFERENCE         S57197
   #authors       Sendai, Y.; Abe, H.; Kikuchi, M.; Satoh, T.; Hoshi, H.
   #journal       Biol. Reprod. (1994) 50:927-934
   #title         Purification and molecular cloning of bovine oviduct-specific
                    glycoprotein.
   #accession     S57197
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-537 ##label SEN
      ##cross-references EMBL:D16639
SUMMARY           #length 537  #checksum 1695

  Query Match            39.1%;  Score 1345;  DB 14;  Length 537;
  Best Local Similarity  50.4%;  Pred. No. 2.27e-230;
  Matches   191;  Conservative   80;  Mismatches 101;  Indels   7;  Gaps   6;

Db      6 glllvlkhhdgaahklvcyftnwafsrpgpasilprdldpflcthlvfafasmsnnqivp 65
          |::::|:   |:| ||||||||||  | | | :||:|||| |||||::|||:|:|:|:
Qy      9 GFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLST 68

Db     66 kdpqdekilypefnklkernrglktllsiggwnfgtvrfttmlstfsnrerfvssviall 125
           :  ||  || ||| ||  |  |||||:|||||||| :|| |::| :||: ||:| | :|
Qy     69 TEWNDET-LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFL 127

Db    126 rthgfdgldlfflypglrgspardrwtfvflleellqafkneaqltmrprlllsaavsgd 185
          |  :|||||   ||| :|||| |:  |  |:::|  ||  ||| : : |||||||::
Qy    128 RKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG 187

Db    186 phvvqkayearllgrlldfisvlsydlhgswekvtghnsplfs-l--pgdpks-sa-yam 240
            |: :||   ::: |||:::::||:|||||||||||||:     :| : | :   |:
Qy    188 QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAV 247

Db    241 nywrqlgvppekllmglptygrtfhllkasqnelraqavgpaspgkytkqagflayyeic 300
            | | | |: ||::|:|||||:| |  :|:  : | | |:::|| :||::| |||||:|
Qy    248 QQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVC 307

Db    301 cfvrrakkrwindqyvpyafkgkewvgyddaisfgykaffikrehfggamvwtldlddfr 360
            : | |::| || ||| |: ::|||:||  ||  |  ::|:  :||||||:||||||
Qy    308 SW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFA 366

Db    361 gyfcgtgpfplvhtlnnll 379
          |: |  | :||::||   |
Qy    367 GFSCNQGRYPLIQTLRQEL 385


RESULT    8
ENTRY             A53918     #type complete
TITLE             chitinase (EC 3.2.1.14) precursor - braconid wasp (Chelonus
                    sp.)
ORGANISM          #formal_name Chelonus sp.
DATE              28-Jul-1995 #sequence_revision 28-Jul-1995 #text_change
                    03-May-1996
ACCESSIONS        A53918
REFERENCE         A53918
   #authors       Krishnan, A.; Nair, P.N.; Jones, D.
   #journal       J. Biol. Chem. (1994) 269:20971-20976
   #title         Isolation, cloning, and characterization of new chitinase
                    stored in active form in chitin-lined venom reservoir.
   #accession     A53918
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-483 ##label KRI
      ##cross-references GB:U10422
KEYWORDS          glycosidase; hydrolase
SUMMARY           #length 483  #molecular-weight 52013  #checksum 5938

  Query Match            34.4%;  Score 1185;  DB 12;  Length 483;
  Best Local Similarity  42.3%;  Pred. No. 7.24e-199;
  Matches   159;  Conservative   100;  Mismatches 100;  Indels   17;  Gaps   11;

Db     19 aspnkvvcyfgawsvyrqgngkfdingidptlcthliysfvgvngkdvkvldpwsdlpgn 78
          :|: |:|||| |: ||||:::|  : :||:|||||||:| |:   ::   : |:|
Qy     19 GSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTE-WND--ET 75

Db     79 ldgfgkftslrkknpsvkimvavggwnagsvpfsqmasdqatreafaqnvvkflqqyqfd 138
          |  :  | :|:|:|| :| ::|:|||| |:  |::| :    |::|  : ::||: | ||
Qy     76 L--YQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFD 133

Db    139 gfdidweypaqrggspadvknmvklckalkkafvq--h----d-yilsaavaapetsask 191
          |:|:|||||: :|::: |      |   | :|| |  :    :  :|||||:| :|
Qy    134 GLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDA 193

Db    192 sydiaemsqyldfinlmtydfhgpwdghtgmhappsasshdsgnelklnvkaavkywlqn 251
          :|::  ::| |||:|||:|||||:|:  || ::|      :||    ||| |||  |||:
Qy    194 GYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQK 253

Db    252 gvpkeklvvgvpaygksftlsnpsnkglgapvsgagtagpytgengllgyneicemqkag 311
          | |  ||::|:|:||:||||:::|:  :||| :|:||:||:| | |:|:| |:|   |::
Qy    254 GTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGA 312

Db    312 dwevvqdnekgvpyavkgnqwvsfddlaaikakaqfikqeglggamvwsietddfkgl-c 370
            : :|| :| ||| : ||||:|||: ::|:|  ::|| |||||||::: ||| |: |
Qy    313 TKQRIQD-QK-VPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSC 370

Db    371 ge-kypvlkalnsvlg 385
           : :||:: :|   |:
Qy    371 NQGRYPLIQTLRQELS 386


RESULT    9
ENTRY             A38221     #type complete
TITLE             chitinase (EC 3.2.1.14) MF1 - nematode (Brugia malayi)
ORGANISM          #formal_name Brugia malayi
DATE              04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
                    03-May-1996
ACCESSIONS        A38221
REFERENCE         A38221
   #authors       Fuhrman, J.A.; Lane, W.S.; Smith, R.F.; Piessens, W.F.;
                    Perler, F.B.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1992) 89:1548-1552
   #title         Transmission-blocking antibodies recognize microfilarial
                    chitinase in brugian lymphatic filariasis.
   #cross-references MUID:92179220
   #accession     A38221
      ##status         preliminary; not compared with conceptual translation
      ##molecule_type nucleic acid; protein
      ##residues       1-504 ##label FUH
      ##cross-references NCBIP:85345
      ##note          sequence extracted from NCBI backbone
KEYWORDS          glycosidase; hydrolase
SUMMARY           #length 504  #molecular-weight 55971  #checksum 7343

  Query Match            31.5%;  Score 1086;  DB 12;  Length 504;
  Best Local Similarity  42.8%;  Pred. No. 1.85e-179;
  Matches   158;  Conservative   94;  Mismatches 102;  Indels   15;  Gaps   11;

Db     27 cyytnwaqyrdgegkflpgnipnglcthilyafakvdelgdskpfewndedtewskgmys 86
          ||:|||||||:||::||| ::  :||||::|||| : :   |   |||| |     :|
Qy     26 CYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTT-EWNDE-T-----LYQ 78

Db     87 avtklretnpglkvllsyggynfgsaiftgiaksaqkterfiksaiaflrknnfdgfdld 146
              |:  || || ||: || |||:  || :  :| : : |::||| |||| :|||:|||
Qy     79 EFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 138
```

```
Db    147 weyp-v-gv-a-e-eh-aklveamktafveeaktsgkqrllltaavsagkgtidgsynve 200
          ||||   |  | : |: : ||: :  || :|| ||||:||||:|||:||   :|::|:|:
Qy    139 WEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVD 198

Db    201 slgknfdllflmsydlhgsweknvdlhgk1hptkgevsgigifntefaadywaskgmpke 260
           :: |:|:: ||:||:|||||     :: |    | :: : :| : | : |  || |
Qy    199 KIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPAS 258

Db    261 kiiigipmyaqgwtldnpsetaigaaasrpssasktnpaggtasyweickylkeggketv 320
          |:|:|:| |::: || ::|:|: :||:|: :::::     ||  :| |:|   | : |: :
Qy    259 KLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRI 317

Db    321 hqegvgaymvkgdqwygydneetirikmkw1kekgyggafiwaldfddftgkscgkgpyp 380
          ::: | :|: : :|| |:|: |::: |:  ||:|| ||| :||||:|||:| ||  | ||
Qy    318 QDQKV-PYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYP 376

Db    381 llnaissel 389
          |: ::  ||
Qy    377 LIQTLRQEL 385


RESULT   10
ENTRY            A56596     #type complete
TITLE            chitinase (EC 3.2.1.14) - tobacco hornworm
ORGANISM         #formal_name Manduca sexta #common_name tobacco hornworm
DATE             11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change
                   03-May-1996
ACCESSIONS       A56596
REFERENCE        A56596
   #authors      Kramer, K.J.; Corpuz, L.; Choi, H.K.; Muthukrishnan, S.
   #journal      Insect Biochem. Mol. Biol. (1993) 23:691-701
   #title        Sequence of a cDNA and expression of the gene encoding
                   epidermal and gut chitinases of Manduca sexta.
   #cross-references MUID:93357793
   #accession    A56596
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-554 ##label KRA
      ##cross-references NCBIN:136417; NCBIP:136418
      ##experimental_source larvae
      ##note          sequence extracted from NCBI backbone
KEYWORDS         glycosidase; hydrolase
SUMMARY          #length 554  #molecular-weight 62203  #checksum 4659

  Query Match           29.9%;  Score 1031;  DB 12;  Length 554;
  Best Local Similarity  38.8%;  Pred. No. 1.04e-168;
  Matches   162;  Conservative  105;  Mismatches 126;  Indels  25;  Gaps  19;

Db     17 vqsdsrarivcyfsnwavyrpgvgrygiedipvekcthiiysfigvtegnsevliidpel 76
          :   | |::||||:||| || | :|:   |:    |||:||:| |:|: : : |    |
Qy     16 IPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTN-H-Q-LST-TEW 71

Db     77 dvdkngfrnftslrsshpsvkfmvavggwaegsskyshmvaqkstrmsfirsvvsflkky 136
          : |   :::| :|:  :| :| ::|:|||  |: |:: ||| : | :|: | : ||:||
Qy     72 N-DETLYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKY 130

Db    137 dfdgld1dweypgaadrggsfsdkdkflylvqelrraf---irv-gkgwel-taavplan 191
           ||||||||||||:  :|::  ||::|  |||:|  ||    :  || : | :|||| :
Qy    131 SFDGLDLDWEYPGS--QGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQ 188

Db    192 frlmegyhvpelcqeldaihvmsydlrgnwagfadvhsplykrphdqwayeklnvndglh 251
            :  || |  : |:|| :::|:||::|:|   :  :|||||  :  |   |||: :::
Qy    189 TYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQ 248

Db    252 1weekgcpsnklvvgipfygrsftlsagnnnyglgtfinkeagggdpapytnatgfwayy 311
          | :|| |::||::|:| ||||||:::::   :|:     :|:| |:|:|:  |  |||
Qy    249 QWLQKGTPASKLILGMPTYGRSFTLASSSDT-RVGA--P-ATGSGTPGPFTKEGGMLAYY 304

Db    312 eictevdkddsgwtkkwdeqgkcpyaykgtqwvgyedprsveikmnwikqkgylgamtwa 371
          |:|:   |  :  || : :: | || ::  ||||::|  |   |:: :|||  ||| ||
Qy    305 EVCS--WKG-A--TKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWA 359

Db    372 idmddfqgl-cge-knplikilhkhms-sytvppphtenttptpewarppstpsdpse 426
          :|:||| |: | : : ||| |:  :| :| :|:   |   | |  :  |   : |::
Qy    360 LDLDDFAGFSCNQGRYPLIQTLRQELSLPY-LPSGTPELEVPKPGQPSEPEHGPSPGQ 416


RESULT   11
ENTRY            A38368     #type complete
TITLE            chitinase (EC 3.2.1.14) precursor - Bacillus circulans
ORGANISM         #formal_name Bacillus circulans
DATE             28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change
                   03-May-1996
ACCESSIONS       A38368
REFERENCE        A38368
   #authors      Watanabe, T.; Suzuki, K.; Oyanagi, W.; Ohnishi, K.; Tanaka,
                   H.
   #journal      J. Biol. Chem. (1990) 265:15659-15665
   #title        Gene cloning of chitinase A1 from Bacillus circulans WL-12
                   revealed its evolutionary relationship to Serratia
                   chitinase and to the type III homology units of
                   fibronectin.
   #cross-references MUID:90368776
   #accession    A38368
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-699 ##label WAT
      ##cross-references GB:J05599
CLASSIFICATION   #superfamily fibronectin type III repeat homology
KEYWORDS         glycosidase; hydrolase
SUMMARY          #length 699  #molecular-weight 73677  #checksum 6757

  Query Match           20.3%;  Score 698;  DB 10;  Length 699;
  Best Local Similarity  40.1%;  Pred. No. 2.46e-104;
  Matches   119;  Conservative   73;  Mismatches  85;  Indels  20;  Gaps  18;

Db    144 qlnk1kqtnpn1ktiisvggwtw-snrfsdvaataatrevfansavdf1rkynfdgvdld 202
          ::| ||  ||:|||:::::|||   : :|:|: |||  |: | |||: |||||:|||:|||
Qy     79 EFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 138

Db    203 weypvsggldgnskrpedkqnytlllskirekldaagavdgk-kylltiasgasatya-a 260
          ||||  |: :| |  : ||: :| |:  : : ::  :   || : ||: |  |: ||  |
Qy    139 WEYP--GS-QG-S-PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDA 193

Db    261 ntelakiaaivdwinimtydfngawqkisahnaplnydpaasaagvpdantfnvaagaqg 320
            |: |||  :| :|:|:|||:|:|:|:::||:|| |      :| : | ::|| |: |
Qy    194 GYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPL-YKRQE-ESGAA-A-SLNVDAAVQQ 249

Db    321 hldagvpaak1v1gvpfygrgwd-gcaqagn-gqyqtctggssvgtweagsfdfydlean 378
          |: | ||:||:||:| |||:   : :      |  | :| ::  | |:| : :|::  :
Qy    250 WLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEV-CS 308

Db    379 yinkngytrywndtakvpylynasnkrfisyddaesvgyktayikskglggamfwel 435
            | |: : :  ||||::   | : :::|| ||   | :|:| |||||| | |
Qy    309 W--K-GATKQRIQDQKVPYIFRD-NQW-VGFDDVESFKTKVSYLKQKGLGGAMVWAL 360


RESULT   12
ENTRY            S51591     #type complete
TITLE            chitinase (EC 3.2.1.14)/lysozyme (EC 3.2.1.17) PZ precursor,
                   pathogenesis-related - common tobacco
ORGANISM         #formal_name Nicotiana tabacum #common_name common tobacco
DATE             15-Jul-1995 #sequence_revision 08-Sep-1995 #text_change
                   03-May-1996
ACCESSIONS       S51591; S51632; S43119
REFERENCE        S51591
   #authors      Heitz, T.; Segond, S.; Kauffmann, S.; Geoffroy, P.; Prasad,
                   V.; Brunner, F.; Fritig, B.; Legrand, M.
   #journal      Mol. Gen. Genet. (1994) 245:246-254
   #title        Molecular characterization of a novel tobacco
                   pathogenesis-related (PR) protein: a new plant
                   chitinase/lysozyme.
   #accession    S51591
      ##molecule_type mRNA
      ##residues      1-378 ##label HEI
      ##cross-references EMBL:X78325
```

```
     ##experimental_source cv. Samsun NN
   #accession    S51632
     ##molecule_type protein
     ##residues      31-36;87-112;252-275;282-305;337-371 ##label HEW
CLASSIFICATION   #superfamily Streptomyces chitinase chi40
KEYWORDS         glycosidase; hydrolase
FEATURE
   1-25               #domain signal sequence #status predicted #label SIG\
   26-371             #product chitinase/lysozyme PZ #status predicted #label
                        MAT\
   371-378            #domain carboxyl-terminal propeptide #status
                        experimental #label PRO
SUMMARY          #length 378  #molecular-weight 42019  #checksum 6689

  Query Match            17.3%;  Score 594;  DB 5;  Length 378;
  Best Local Similarity  32.7%;  Pred. No. 1.33e-84;
  Matches   106;  Conservative   87;  Mismatches 108;  Indels  23;  Gaps  18;

Db      43 nnidstlfthlfcafadlnpqsnqliispenqds-fsqftstvqrknpsvktflsiaggr 101
           :::|::| |||: ||| :   ::||  :  |::: : :| : : ::|| :|||:|:| ||:
Qy      44 KDLDPSLCTHLIYAFAGMT--NHQLSTTEWNDETLYQEFNG-LKKMNPKLKTLLAI-GGW 99

Db     102 a-dttaygimarqpnsrksfidssirlarqfgfhqidldweyp--l-s-atdmtnlgill 156
             |  :  |   :|:| :|::|:||: | ::| |||||||||    | | |:  :  |:
Qy     100 NFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLV 159

Db     157 newrtainmearnsgraallltaavsysprvnglnypvesvarnlnwinlmaydfygpnw 216
           :   |:  ||: ||:  |||:|||: :       | |: :|:||: :|||||| |: |
Qy     160 QDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGS-W 218

Db     217 spsqtnshaqlfdpvnh--isgs---dg-inawiqagvptkklvlgipfygyawrlvnpn 270
             |  :: |:   :    ::|   |: :  |:| | |: ||:||:| || :  | :::
Qy     219 EKV-TGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSS 277

Db     271 ihdlrapaagksnvga-vddgsm-tynrirdyivqsrattvynativgdycysgsnwisy 328
           : |||:| :  |:   :|:| :|  :      ::   : :   |  | :  : |:::
Qy     278 DTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQR-IQDQK-VP-YIFRDNQWVGF 334

Db     329 ddtqsvrnkvnyvkgrgllgyfaw 352
           || :| : ||:|:| :|| |   |
Qy     335 DDVESFKTKVSYLKQKGLGGAMVW 358


RESULT   13
ENTRY            S47133     #type complete
TITLE            chitinase (EC 3.2.1.14) - fungus (Trichoderma harzianum)
ORGANISM         #formal_name Trichoderma harzianum
DATE             23-Nov-1994 #sequence_revision 03-Aug-1995 #text_change
                   10-May-1996
ACCESSIONS       S47133
REFERENCE        S47133
   #authors      Carsolio, C.; Gutierrez, A.; Jimenez, B.; van Montagu, M.;
                   Herrera-Estrella, A.
   #submission   submitted to the EMBL Data Library, June 1994
   #description  Characterization of ech-42, a Trichoderma harzianum
                   endochitinase gene expressed during mycoparasitism.
   #accession    S47133
     ##status         preliminary
     ##molecule_type DNA
     ##residues      1-424 ##label CAR
     ##cross-references EMBL:X79381
GENETICS
   #introns      48/2; 81/2; 98/1
CLASSIFICATION   #superfamily Streptomyces chitinase chi40
KEYWORDS         glycosidase; hydrolase
SUMMARY          #length 424  #molecular-weight 46298  #checksum 8773

  Query Match            16.9%;  Score 582;  DB 5;  Length 424;
  Best Local Similarity  36.0%;  Pred. No. 2.42e-82;
  Matches   102;  Conservative   69;  Mismatches  95;  Indels  17;  Gaps  13;

Db     112 qlfk-lkkanrnlkvmlsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidv 169
           | |: ||| | :|| :|:|||| :| :|    :|  ||  | ::|| |::  :|||:|:
Qy      78 QEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 137

Db     170 dweyp-addtqatnmvlllkeirsqldayaaqy-apg-yhfllsiaapagpehy-sflhm 225
           ||||| :: : | ::  :   ::   :|:  :  ::|  ::||| | |||     :   :
Qy     138 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 197

Db     226 sdlgqvldyvnlmaydyagswssysghdanlfanpsnpnssp-yntdqaikdyikggvpa 284
             ::| ||:|||||||: |||   :||:: |:    :: :::  | | |: : :  | ||
Qy     198 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 257

Db     285 skivlgmpiygraf--estgg--igqtysgigs-gsw--engiwdy-kvlpkagatvqyd 336
           ||::|||| |||:|   |::   :|   :| |: |:   | |:  |  | :  ||| |
Qy     258 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQ-- 315

Db     337 svaqayysydpsskelisfdtpdmintkvsylknlglggsmfw 379
           : :   :|    :: ::||  : ::|||||||  ||||:| |
Qy     316 RIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT   14
ENTRY            JQ1975     #type complete
TITLE            chitinase (EC 3.2.1.14) 1 precursor - imperfect fungus
                   (Aphanocladium album)
ORGANISM         #formal_name Aphanocladium album
DATE             03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change
                   03-May-1996
ACCESSIONS       JQ1975
REFERENCE        JQ1975
   #authors      Blaiseau, P.L.; Lafay, J.F.
   #journal      Gene (1992) 120:243-248
   #title        Primary structure of a chitinase-encoding gene (chi1) from
                   the filamentous fungus Aphanocladium album: Similarity to
                   bacterial chitinases.
   #accession    JQ1975
     ##molecule_type mRNA
     ##residues      1-423 ##label BLA
     ##cross-references GB:X64104
COMMENT     This enzyme is essential for the degradation of insoluble chitin.
GENETICS
   #gene         chi1
   #introns      47/2; 80/2; 97/1
CLASSIFICATION   #superfamily Streptomyces chitinase chi40
KEYWORDS         glycosidase; hydrolase
FEATURE
   1-34               #domain signal sequence #status predicted #label SIG\
   35-423             #product chitinase 1 #status predicted #label MAT
SUMMARY          #length 423  #molecular-weight 46072  #checksum 8802

  Query Match            16.7%;  Score 576;  DB 5;  Length 423;
  Best Local Similarity  36.2%;  Pred. No. 3.26e-81;
  Matches   101;  Conservative   65;  Mismatches  95;  Indels  18;  Gaps  16;

Db     115 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 173
           ||| | ::| :|:|||| :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy      83 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 142

Db     174 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 228
            :::: | ::  :   |:   :::  :  : |  ::||| | ||| : |    :  : ::
Qy     143 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 201

Db     229 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 287
             ||::||||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy     202 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 261

Db     288 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 339
           |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||   :  |
Qy     262 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 321

Db     340 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 378
             | :   :: ::||    : |||| || |||||:| |
Qy     322 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358
```

```
RESULT   15
ENTRY            S51369     #type complete
TITLE            chitinase - fungus (Trichoderma harzianum)
ORGANISM         #formal_name Trichoderma harzianum
DATE             15-Jul-1995 #sequence_revision 19-Apr-1996 #text_change
                   03-May-1996
ACCESSIONS       S51369
REFERENCE        S51369
   #authors      Garcia, I.; Lora, J.M.; de la Cruz, J.; Benitez, T.; Llobell,
                   A.; Pintor-Toro, J.A.
   #journal      Curr. Genet. (1994) 27:83-89
   #title        Cloning and characterization of a chitinase (CHIT42) cDNA
                   from the mycoparasitic fungus Trichoderma harzianum.
   #accession    S51369
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-423 ##label GAR
SUMMARY          #length 423  #molecular-weight 46056  #checksum 2000

  Query Match           16.3%;  Score 561;  DB 12;  Length 423;
  Best Local Similarity  35.7%;  Pred. No. 2.14e-78;
  Matches   101;  Conservative   71;  Mismatches  94;  Indels  17;  Gaps  13;

Db    111 qlfk-vkkanrglkvllsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidi 168
          | |: :|| |  || ||:|||  :| :|    :|  ||  | ::|| |::  :|||:|:
Qy     78 QEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 137

Db    169 dweyp-adatqasnmilllkevrsqrdayaaqy-apg-yhflltiaapagkdny-sklrl 224
          ||||| :::: | ::  :   |:   :|:  :  ::|  ::||: | |||     :   :
Qy    138 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 197

Db    225 adlgqvldyinlmaydyagsfspltghdanlfnnpsnpnatp-fntdsavkdyinggvpa 283
            ::| ||::|||||: ||   :|||:: |::   :: |:: :| |:|| : :  | ||
Qy    198 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 257

Db    284 nkivlgmpiygrsf--qntag--igqtyngvgs-gsw--eagiwdykalpk-agatvqyd 335
          :|::|||| |||||   :::   :|    | |: |:   |:|:  |  :    ||| |
Qy    258 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRI 317

Db    336 svakgyysynsatkelisfdtpdmintkvaylkslglggsmfw 378
            | | :    :: ::||  : ::|||:|||  ||||:| |
Qy    318 QDQKVPYIFRD--NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


Search completed: Fri Jun 27 16:51:51 1997
Job time : 159 secs.
```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 27 16:46:47 1997; MasPar time 14.60 Seconds
676.776 Million cell updates/sec
Tabular output not generated.

Title: >US-08-663-618A-2
Description: (1-466) from US08663618A.pep
Perfect Score: 3443
Sequence: 1 MVRSVAWAGFMVLLMIPWGS.....QQSCPTGLVFSNSCKCCTWN 466

Scoring table: PAM 150
Gap 11

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot34
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11

Statistics: Mean 50.012; Variance 83.709; scale 0.597

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1519	44.1	383	4	GP39_HUMAN	CARTILAGE GLYCOPROTEI	0.00e+00
2	1086	31.5	504	2	CHIT_BRUMA	ENDOCHITINASE PRECURS	2.83e-227
3	1031	29.9	554	2	CHIT_MANSE	ENDOCHITINASE PRECURS	1.18e-213
4	698	20.3	699	2	CHI1_BACCI	CHITINASE A1 PRECURSO	3.91e-132
5	576	16.7	423	2	CHI1_APHAL	CHITINASE 1 PRECURSOR	7.00e-103
6	561	16.3	423	2	CHI4_TRIHA	42 KD ENDOCHITINASE P	2.55e-99
7	546	15.9	427	2	CHI1_COCIM	ENDOCHITINASE 1 PRECU	9.12e-96
8	437	12.7	563	2	CHIA_SERMA	CHITINASE A PRECURSOR	2.88e-70
9	395	11.5	551	2	CHIT_NPVAC	PROBABLE ENDOCHITINAS	1.25e-60
10	345	10.0	820	2	CHIA_ALTSO	CHITINASE A PRECURSOR	2.40e-49
11	337	9.8	499	2	CHIB_SERMA	CHITINASE B PRECURSOR	1.46e-47
12	301	8.7	619	2	CHIT_STRLI	CHITINASE C PRECURSOR	1.25e-39
13	287	8.3	610	2	CHIT_STRPL	CHITINASE 63 PRECURSO	1.37e-36
14	263	7.6	1146	5	KTXA_KLULA	KILLER TOXIN ALPHA AN	1.89e-31
15	172	5.0	36	4	GP39_BOVIN	CARTILAGE GLYCOPROTEI	4.12e-13
16	172	5.0	83	3	EDOS_PAPAN	ESTRADIOL-DEPENDENT O	4.12e-13
17	154	4.5	541	11	YKC5_CAEEL	HYPOTHETICAL 55.6 KD	8.28e-10
18	121	3.5	597	2	CHIX_STROI	EXOCHITINASE 1 PRECUR	3.24e-04
19	110	3.2	497	2	CHID_BACCI	CHITINASE D PRECURSOR	1.57e-02
20	108	3.1	452	4	HEMF_RHOSH	COPROPORPHYRINOGEN II	3.10e-02
21	104	3.0	422	10	UMUC_ECOLI	UMUC PROTEIN.	1.17e-01
22	99	2.9	289	11	YFK2_YEAST	HYPOTHETICAL 31.9 KD	5.83e-01
23	100	2.9	364	11	YBR5_YEAST	HYPOTHETICAL 40.9 KD	4.25e-01
24	99	2.9	398	9	SPEB_STRPY	EXOTOXIN TYPE B PRECU	5.83e-01
25	101	2.9	958	1	AMYG_SCHOC	GLUCOAMYLASE 1 PRECUR	3.09e-01
26	101	2.9	1045	7	PRTT_SERMA	EXTRACELLULAR SERINE	3.09e-01
27	101	2.9	1070	1	AGLU_CANTS	ALPHA-GLUCOSIDASE PRE	3.09e-01
28	96	2.8	297	5	HMX1_HUMAN	HOMEOBOX PROTEIN MSX-	1.48e+00
29	95	2.8	353	11	YQHT_BACSU	PUTATIVE PEPTIDASE IN	2.00e+00
30	98	2.8	492	9	SYK_THETH	LYSYL-TRNA SYNTHETASE	7.97e-01
31	95	2.8	492	7	PDI_SCHPO	PUTATIVE PROTEIN DISU	2.00e+00
32	96	2.8	554	5	HYES_RAT	SOLUBLE EPOXIDE HYDRO	1.48e+00
33	95	2.8	554	5	HYES_MOUSE	SOLUBLE EPOXIDE HYDRO	2.00e+00
34	97	2.8	563	5	K2CB_HUMAN	KERATIN, TYPE II CYTO	1.09e+00
35	97	2.8	563	5	K2CA_HUMAN	KERATIN, TYPE II CYTO	1.09e+00
36	97	2.8	563	5	K2CC_HUMAN	KERATIN, TYPE II CYTO	1.09e+00
37	95	2.8	563	5	K2CE_HUMAN	KERATIN, TYPE II CYTO	2.00e+00
38	95	2.8	563	5	K2CF_HUMAN	KERATIN, TYPE II CYTO	2.00e+00
39	98	2.8	1045	7	PRTS_SERMA	EXTRACELLULAR SERINE	7.97e-01
40	95	2.8	1333	10	VGR1_MOUSE	VASCULAR ENDOTHELIAL	2.00e+00
41	95	2.8	1409	2	COPI_DROME	COPIA PROTEIN.	2.00e+00
42	94	2.7	344	4	GCH2_PICGU	GTP CYCLOHYDROLASE II	2.71e+00
43	94	2.7	404	6	MOEA_HAEIN	MOLYBDOPTERIN BIOSYNT	2.71e+00
44	94	2.7	789	10	VIB4_AGRT5	VIRB4 PROTEIN PRECURS	2.71e+00
45	94	2.7	993	11	YAJ1_SCHPO	PUTATIVE FAMILY 31 GL	2.71e+00

ALIGNMENTS

RESULT 1
ID GP39_HUMAN STANDARD; PRT; 383 AA.
AC P36222; P30923;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE CARTILAGE GLYCOPROTEIN-39 PRECURSOR (GP-39) (39 KD SYNOVIAL PROTEIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=CARTILAGE;
RX MEDLINE; 94064658.
RA HAKALA B.E., WHITE C., RECKLIES A.D.;
RL J. BIOL. CHEM. 268:25803-25810(1993).
RN [2]
RP SEQUENCE OF 22-45.
RX MEDLINE; 90328983.
RA NYIRKOS P., GOLDS E.E.;
RL BIOCHEM. J. 269:265-268(1990).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE CAPACITY OF CELLS TO
CC RESPOND TO AND COPE WITH CHANGES IN THEIR ENVIRONMENT.
CC -!- TISSUE SPECIFICITY: PRESENT IN ARTICULAR CHONDROCYTES, SYNOVIAL
CC CELLS AS WELL AS IN LIVER. UNDETECTABLE IN MUSCLE TISSUES, LUNG,
CC PANCREAS, MONONUCLEAR CELLS, OR FIBROBLASTS.
CC -!- PTM: GLYCOSYLATED.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
DR EMBL; M80927; G348912; -.
DR PIR; S10677; S10677.
DR PIR; A33162; A33162.
DR PROSITE; PS01095; CHITINASE_18.
KW GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 21
FT CHAIN 22 383 CARTILAGE GLYCOPROTEIN-39.
FT CARBOHYD 60 60 POTENTIAL.
SQ SEQUENCE 383 AA; 42613 MW; D0A5280F CRC32;

Query Match 44.1%; Score 1519; DB 4; Length 383;
Best Local Similarity 52.8%; Pred. No. 0.00e+00;
Matches 201; Conservative 88; Mismatches 83; Indels 9; Gaps 5;

```
Db      8 tgfvvlvllqccsayklvcyytswsqyregdgscfpdaldrflcthiiysfanisndhid 67
          :||:||::   || |||||:|:|:|||:|::  :|  ||  ||||:||:|| ::| ::
Qy      8 AGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLS 67
```

```
Db      68 twewndvtlygmlntlknrnpnlktllsvggwnfgsqrfskiasntqsrrtfiksvppfl 127
Qy      68 TTEWNDETLYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFL 127

Db     128 rthgfdgldlawlypgrr-----dkqhfttlikemkaefikeaq-pgkkqlllsaalsag 181
Qy     128 RKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG 187

Db     182 kvtidssydiakisqhldfisimtydfhgawrgttghhsplfrgqedaspdrfsntdyav 241
Qy     188 QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAV 247

Db     242 gymlrlgapasklvmgiptfgrsftlass-etgvgapisgpgipgrftkeagtlayyeic 300
Qy     248 QQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVC 307

Db     301 dflrgatvhrtlgqqvpyatkghqwvgyddqesvkskvqylkdrqlagamvwaldlddfq 360
Qy     308 SW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFA 366

Db     361 gsfcgqdlrfpltnaikdala 381
Qy     367 GFSCNQG-RYPLIQTLRQELS 386
```

RESULT 2
ID CHIT_BRUMA STANDARD; PRT; 504 AA.
AC P29030;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE ENDOCHITINASE PRECURSOR (EC 3.2.1.14) (MF1 ANTIGEN).
OS BRUGIA MALAYI.
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; SPIRURIDA.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 92179220.
RA FUHRMAN J.A., LANE W.S., SMITH R.F., PIESSENS W.F., PERLER F.B.;
RL PROC. NATL. ACAD. SCI. U.S.A. 89:1548-1552(1992).
CC -!- FUNCTION: THE MF1 ANTIGEN IS A MICROFILARIAL CHITINASE, WHICH MAY
CC FUNCTION TO DEGRADE CHITIN-CONTAINING STRUCTURES IN THE MICRO-
CC FILARIA OR IN ITS MOSQUITO VECTOR DURING PARASITE DEVELOPMENT AND
CC TRANSMISSION.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC -!- PTM: O-GLYCOSYLATED.
CC -!- KNOWN TO BIND CALCIUM.
CC -!- DEVELOPMENTAL STAGE: THE APPEARANCE OF THE MF1 ANTIGEN CORRESPOND
CC WITH THE ONSET OF THE PARASITE'S ABILITY TO INFECT THE MOSQUITO.
CC -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; M73689; G156064; -.
DR PIR; A38221; A38221.
DR PROSITE; PS01095; CHITINASE_18.
KW HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ANTIGEN;
KW REPEAT; GLYCOPROTEIN; CALCIUM-BINDING.
FT SIGNAL 1 22
FT CHAIN 23 504 ENDOCHITINASE.
FT DOMAIN 23 400 CATALYTIC.
FT DOMAIN 401 450 SER/THR-RICH (LINKER).
FT DOMAIN 407 448 3 X 14 AA APPROXIMATE TANDEM REPEATS.
FT ACT_SITE 148 148 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 504 AA; 55971 MW; 4DA7E5EB CRC32;

Query Match 31.5%; Score 1086; DB 2; Length 504;
Best Local Similarity 42.8%; Pred. No. 2.83e-227;
Matches 158; Conservative 94; Mismatches 102; Indels 15; Gaps 11;

```
Db      27 cyytnwaqyrdgegkflpgnipnglcthilyafakvdelgdskpfewndedtewskgmys 86
Qy      26 CYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTT-EWNDE-T-----LYQ 78

Db      87 avtklretnpglkvllsyggynfgsaiftgiaksaqkterfiksaiaflrknnfdgfdld 146
Qy      79 EFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 138

Db     147 weyp-v-gv-a-e-eh-aklveamktafveeaktsgkqrllltaavsagkgtidgsynve 200
Qy     139 WEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVD 198

Db     201 slgknfdllflmsydlhgsweknvdlhgklhptkgevsgigifntefaadywaskgmpke 260
Qy     199 KIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPAS 258

Db     261 kiiigipmyaqgwtldnpsetaigaaasrpssasktnpaggtasyweickylkeggketv 320
Qy     259 KLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRI 317

Db     321 hqegvgaymvkgdqwygydneetirikmkwlkekgyggafiwaldfddftgkscgkgpyp 380
Qy     318 QDQKV-PYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYP 376

Db     381 llnaissel 389
Qy     377 LIQTLRQEL 385
```

RESULT 3
ID CHIT_MANSE STANDARD; PRT; 554 AA.
AC P36362;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
OS MANDUCA SEXTA (TOBACCO HAWKMOTH) (TOBACCO HORNWORM).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; LEPIDOPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93357793.
RA KRAMER K.J., CORPUZ L., CHOI H.K., MUTHUKRISHNAN S.;
RL INSECT BIOCHEM. MOL. BIOL. 23:691-701(1993).
CC -!- FUNCTION: DIGEST CHITIN IN THE EXOSKELETON DURING THE MOLTING
CC PROCESS.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC -!- DEVELOPMENTAL STAGE: HIGH LEVELS SEEN IN THE EPIDERMIS ON DAY 0,
CC BUT RAPIDLY DISAPPEARS AND IS UNDETECTED ON DAYS 1-4 OF FIFTH
CC INSTAR. IT REAPPEARS ON DAY 5 AND PEAKS ON DAY 7 AFTER WHICH A
CC RAPID DECLINE IS SEEN. IN THE GUT IS DETECTED ON DAY 6 WITH LOWER
CC LEVELS SEEN ON DAYS 0, 7 AND 8.
CC -!- TISSUE SPECIFICITY: EPIDERMIS AND GUT.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; U02270; G406049; -.
DR EMBL; S64757; -; NOT_ANNOTATED_CDS.
DR HSSP; P07254; 1CTN.
DR PROSITE; PS01095; CHITINASE_18.
KW HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 554 ENDOCHITINASE.
FT DOMAIN 396 453 SER/THR-RICH.
FT ACT_SITE 146 146 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 85 85 POTENTIAL.
FT CARBOHYD 303 303 POTENTIAL.
FT CARBOHYD 407 407 POTENTIAL.
FT CARBOHYD 545 545 POTENTIAL.
SQ SEQUENCE 554 AA; 62203 MW; FA87F8AD CRC32;

Query Match 29.9%; Score 1031; DB 2; Length 554;
Best Local Similarity 38.8%; Pred. No. 1.18e-213;
Matches 162; Conservative 105; Mismatches 126; Indels 25; Gaps 19;

```
Db      17 vqsdsrarivcyfsnwavyrpgvgrygiedipvekcthiiysfigvtegnsevliidpel 76
```

```
Qy      16 IPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTN-H-Q-LST-TEW 71

Db      77 dvdkngfrnftslrsshpsvkfmvavggwaegsskyshmvaqkstrmsfirsvvsflkky 136
           : |   :::| :|:  :| :| ::|:|||  |: |:: |||  : | :|: | : ||:||
Qy      72 N-DETLYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKY 130

Db     137 dfdgldldweypgaadrggsfsdkdkflylvqelrraf---irv-gkgwel-taavplan 191
           ||||||||||||:  :|::  ||::|  |||:|  ||    :  || : | :|||| :
Qy     131 SFDGLDLDWEYPGS--QGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQ 188

Db     192 frlmegyhvpelcqeldaihvmsydlrgnwagfadvhsplykrphdqwayeklnvndglh 251
             :  || |  : |:|| :::|:||::|:|   :  :|||||  :  |   |||: ::: 
Qy     189 TYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQ 248

Db     252 lweekgcpsnklvvgipfygrsftlsagnnnyglgtfinkeagggdpapytnatgfwayy 311
           | :|| |::||::|:| |||||||:::::   :|:     :|:| |:|:|:  |  |||
Qy     249 QWLQKGTPASKLILGMPTYGRSFTLASSSDT-RVGA--P-ATGSGTPGPFTKEGGMLAYY 304

Db     312 eictevdkddsgwtkkwdeqgkcpyaykgtqwvgyedprsveikmnwikqkgylgamtwa 371
           |:|:   |  :  || : :: | || ::  ||||::|  |   |:: :|||  ||| ||
Qy     305 EVCS--WKG-A--TKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWA 359

Db     372 idmddfqgl-cge-knplikilhkhms-sytvppphtenttptpewarppstpsdpse 426
           :|:||| |: | : : ||| |:  :| :| :|:   |   | |  :  |   : |::
Qy     360 LDLDDFAGFSCNQGRYPLIQTLRQELSLPY-LPSGTPELEVPKPGQPSEPEHGPSPGQ 416


RESULT    4
ID   CHI1_BACCI     STANDARD;      PRT;   699 AA.
AC   P20533;
DT   01-FEB-1991 (REL. 17, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE A1 PRECURSOR (EC 3.2.1.14).
GN   CHIA1.
OS   BACILLUS CIRCULANS.
OC   PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WL-12;
RX   MEDLINE; 90368776.
RA   WATANABE T., SUZUKI K., OYANAGI W., OHNISHI K., TANAKA H.;
RL   J. BIOL. CHEM. 265:15659-15665(1990).
RN   [2]
RP   MUTAGENESIS.
RC   STRAIN=WL-12;
RX   MEDLINE; 93366760.
RA   WATANABE T., KOHORI K., MIYASHITA K., FUJII T., SAKAI H.,
RA   UCHIDA M., TANAKA H.;
RL   J. BIOL. CHEM. 268:18567-18572(1993).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; M57601; G142688; -.
DR   PIR; A38368; A38368.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; REPEAT.
FT   SIGNAL        1     41
FT   CHAIN        42    699       CHITINASE A1.
FT   DOMAIN       42    460       CATALYTIC.
FT   DOMAIN      465    549       FIBRONECTIN TYPE-III (R-1).
FT   DOMAIN      560    644       FIBRONECTIN TYPE-III (R-2).
FT   ACT_SITE    204    204       PROTON DONOR (PROBABLE).
FT   MUTAGEN     200    200       D->N: DECREASE IN ACTIVITY.
FT   MUTAGEN     200    200       D->E: NO CHANGE IN ACTIVITY.
FT   MUTAGEN     204    204       E->D,Q: LOSS OF ACTIVITY.
SQ   SEQUENCE   699 AA;  73677 MW;  5A1777CC CRC32;

  Query Match           20.3%;  Score 698;  DB 2;  Length 699;
  Best Local Similarity  40.1%;  Pred. No. 3.91e-132;
  Matches   119;  Conservative   73;  Mismatches  85;  Indels  20;  Gaps  18;

Db     144 qlnklkqtnpnlktiisvggwtw-snrfsdvaataatrevfansavdflrkynfdgvdld 202
           ::| ||  ||:|||::::|||   : :|:|: |||  |: | |||: |||||:|||:|||
Qy      79 EFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 138

Db     203 weypvsggldgnskrpedkqnytlllskirekldaagavdgk-kylltiasgasatya-a 260
           ||||  |: :| |  : ||: :| |:  : : ::   :   || : ||: |  |: ||  |
Qy     139 WEYP--GS-QG-S-PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDA 193

Db     261 ntelakiaaivdwinimtydfngawqkisahnaplnydpaasaagvpdantfnvaagaqg 320
             |: |||  :| :|:|:|||:|:|:|:::||:|| |      :| : | ::|| |: |
Qy     194 GYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPL-YKRQE-ESGAA-A-SLNVDAAVQQ 249

Db     321 hldagvpaaklvlgvpfygrgwd-gcaqagn-gqyqtctggssvgtweagsfdfydlean 378
           |: | ||:||:|:|:| |||:   : :     |   | :| ::  | |:| : :|::  :
Qy     250 WLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEV-CS 308

Db     379 yinkngytrywndtakvpylynasnkrfisyddaesvgyktayikskglggamfwel 435
           | | |: : :  ||||::   | : :::|| ||   | :|:| |||||||  | |
Qy     309 W--K-GATKQRIQDQKVPYIFRD-NQW-VGFDDVESFKTKVSYLKQKGLGGAMVWAL 360


RESULT    5
ID   CHI1_APHAL     STANDARD;      PRT;   423 AA.
AC   P32470;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   CHITINASE 1 PRECURSOR (EC 3.2.1.14).
GN   CHI1.
OS   APHANOCLADIUM ALBUM.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ETHM 483;
RX   MEDLINE; 93013040.
RA   BLAISEAU P.-L., LAFAY J.-F.;
RL   GENE 120:243-248(1992).
RN   [2]
RP   SEQUENCE OF 35-57.
RX   MEDLINE; 92136437.
RA   BLAISEAU P.-L., KUNZ C., GRISON R., BERTHEAU Y., BRYGOO Y.;
RL   CURR. GENET. 21:61-66(1992).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; X64104; G429026; -.
DR   PIR; JQ1975; JQ1975.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ZYMOGEN.
FT   SIGNAL        1     22       POTENTIAL.
FT   PROPEP       23     34       POTENTIAL.
FT   CHAIN        35    423       CHITINASE 1.
FT   ACT_SITE    171    171       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   423 AA;  46058 MW;  88456E6E CRC32;

  Query Match           16.7%;  Score 576;  DB 2;  Length 423;
  Best Local Similarity  36.2%;  Pred. No. 7.00e-103;
  Matches   101;  Conservative   65;  Mismatches  95;  Indels  18;  Gaps  16;

Db     115 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 173
           ||| | ::| :|:|||| :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy      83 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 142

Db     174 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 228
            :::: | ::  :  |:  :::  :  : |  ::||| | ||| : |    :  : ::
Qy     143 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 201

Db     229 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 287
```

```
          ||::||||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    202 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 261

Db    288 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 339
          |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    262 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 321

Db    340 gcysydpstkelisfdtpamistkvswlkgkglggtmfw 378
            | :    :: ::||    : |||| || |||||:| |
Qy    322 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT    6
ID   CHI4_TRIHA     STANDARD;      PRT;   423 AA.
AC   P48827;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   42 KD ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
GN   CHIT42.
OS   TRICHODERMA HARZIANUM.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 35-52; 93-107; 371-385 & 397-414.
RX   MEDLINE; 95269313.
RA   GARCIA I., LORA J.M., LA CRUZ J., BENITEZ T., LLOBELL A.,
RA   PINTOR-TORO J.A.;
RL   CURR. GENET. 27:83-89(1994).
CC   -!- FUNCTION: MORPHOGENETIC ROLE DURING APICAL GROWTH,CELL DIVISION
CC       AND DIFFERENTIATION (CELL WALL MORPHOGENESIS). ANTIFUNGAL
CC       AGENT.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- INDUCTION: SPECIFICALLY INDUCED BY CHITIN AND IS CATABOLITE
CC       REPRESSED.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; S78423; G999376; -.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ZYMOGEN;
KW   CHITIN-BINDING.
FT   SIGNAL        1     22       POTENTIAL.
FT   PROPEP       23     34
FT   CHAIN        35    423       42 KD ENDOCHITINASE.
FT   ACT_SITE    171    171       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    218    218       POTENTIAL.
SQ   SEQUENCE   423 AA;  46056 MW;  899DA50A CRC32;

  Query Match           16.3%;  Score 561;  DB 2;  Length 423;
  Best Local Similarity  35.7%;  Pred. No. 2.55e-99;
  Matches   101;  Conservative   71;  Mismatches  94;  Indels  17;  Gaps  13;

Db    111 qlfk-vkkanrglkvllsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidi 168
          | |: :|| |  || ||:|||| :| :|    :|  ||  | ::|| |::  :|||:|:
Qy     78 QEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 137

Db    169 dweyp-adatqasnmilllkevrsqrdayaaqy-apg-yhflltiaapagkdny-sklrl 224
          ||||| :::: | ::  :   |:   :|:  :  ::|  ::||: | |||    :   :
Qy    138 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 197

Db    225 adlgqvldyinlmaydyagsfspltghdanlfnnpsnpnatp-fntdsavkdyinggvpa 283
            ::| ||::|||||: ||   :|||:: |::   :: |:: :| |:|| : :  | ||
Qy    198 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 257

Db    284 nkivlgmpiygrsf--qntag--igqtyngvgs-gsw--eagiwdykalpk-agatvqyd 335
          :|::|||| |||||   :::   :|    | |: |:   |:|:  |  :    ||| |
Qy    258 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRI 317

Db    336 svakgyysynsatkelisfdtpdmintkvaylkslglggsmfw 378
            |  | :    :: ::||  : ::|||:|||  ||||:| |
Qy    318 QDQKVPYIFRD--NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT    7
ID   CHI1_COCIM     STANDARD;      PRT;   427 AA.
AC   P54196;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   ENDOCHITINASE 1 PRECURSOR (EC 3.2.1.14) (COMPLEMENT-FIXATION ANTIGEN)
DE   (CF-ANTIGEN) (CF-AG).
GN   CTS1.
OS   COCCIDIOIDES IMMITIS.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C735;
RX   MEDLINE; 96144270.
RA   PISHKO E.J., KIRKLAND T.N., COLE G.T.;
RL   GENE 167:173-177(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SILVEIRA;
RA   YANG C., ZHU Y., MAGEE D.M., COX R.A.;
RL   SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; L41663; G1200190; -.
DR   EMBL; U51271; G1256769; -.
DR   EMBL; U33265; G1255728; -.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL;
KW   GLYCOPROTEIN.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?    427       ENDOCHITINASE 1.
FT   CARBOHYD    387    387       POTENTIAL.
FT   CONFLICT     15     47       RWLSLRCLFCELGRFMFTLSTVTAVTVVTDDIQ -> VQAS
FT                                SMSSMPNYYPVPEAPEGGFRSVVYFVNW (IN REF. 2).
FT   CONFLICT    199    199       K -> N (IN REF. 2).
SQ   SEQUENCE   427 AA;  47629 MW;  6CB9AF73 CRC32;

  Query Match           15.9%;  Score 546;  DB 2;  Length 427;
  Best Local Similarity  34.1%;  Pred. No. 9.12e-96;
  Matches    95;  Conservative   71;  Mismatches  94;  Indels  19;  Gaps  17;

Db    115 lkknnrnlktllsiggwty-spnfktpasteegrkkfadtslklmkdlgfdgididweyp 173
          ||| | :|||||:|||| : : :|    :| : |  | ::::::::  :|||:|:|||||
Qy     83 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 142

Db    174 ede-kqandfvlllkacrealdaysak-hpngk-kflltiaspagpqny-nk-lklaemd 228
            :   | |   :    ::  :|:    : :|| ::||: | ||| | | :    :  :
Qy    143 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 201

Db    229 kyldfwnlmaydfsgswdkvsghmsnvfpsttkp-estpfssdkavkdyikagvpankiv 287
            ||| ||||||| |||:||:|| | ::      :  ::::: | || : :  | ||:|::
Qy    202 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 261

Db    288 lgmplygraf--ast-dg-igtsfngvgg-gsw--engvwdykdmpq-qgaqvteledia 339
          |||| |||:|  ||: |  :|::  | |  |:   | |:  | ::    ||    ::|
Qy    262 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 321

Db    340 asysydknkrylisydtvkiagkkaeyitkngmgggm-w 377
            :| : ::::  :::| |     |  |:  :|:||:| |
Qy    322 VPYIF-RDNQW-VGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT    8
ID   CHIA_SERMA     STANDARD;      PRT;   563 AA.
AC   P07254;
DT   01-APR-1988 (REL. 07, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE A PRECURSOR (EC 3.2.1.14).
GN   CHIA.
OS   SERRATIA MARCESCENS.
```

```
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   ENTEROBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KOO J.C., LIM C.O., CHOI Y.J., KIM C.Y., BAHK J.D., LEE S.Y.,
RA   CHO M.J.;
RL   SUBMITTED (JAN-1993) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 990 / QMB1466;
RA   JONES J.D.G., GRADY K.L., SUSLOW T.V., BEDBROOK J.R.;
RL   EMBO J. 5:467-473(1986).
RN   [3]
RP   REVISIONS, AND X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE; 95219379.
RA   PERRAKIS A., TEWS I., DAUTER Z., OPPENHEIM A.B., CHET I., WILSON K.S.,
RA   VORGIAS C.E.;
RL   STRUCTURE 2:1169-1180(1994).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; L01455; G152818; -.
DR   EMBL; X03657; G46831; -.
DR   PIR; A25090; A25090.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL.
FT   SIGNAL        1     23
FT   CHAIN        24    563       CHITINASE A.
FT   DOMAIN      150    563       CATALYTIC.
FT   ACT_SITE    315    315       PROTON DONOR (PROBABLE).
FT   ACT_SITE    391    391       PROBABLE.
FT   CONFLICT     76     77       TA -> GP (IN REF. 2).
FT   CONFLICT    139    139       A -> P (IN REF. 2).
FT   CONFLICT    226    226       V -> I (IN REF. 2).
FT   CONFLICT    395    395       P -> A (IN REF. 1 AND 2).
FT   CONFLICT    410    430       PAWKPDTAYTTVNGVNALLAQ -> RPGSRHRLHHGERRQC
FT                                AAGQ (IN REF.1 AND 2).
FT   CONFLICT    437    437       V -> I (IN REF. 2).
FT   CONFLICT    464    467       ATGP -> HRA (IN REF. 2).
FT   CONFLICT    473    473       K -> E (IN REF. 2).
FT   CONFLICT    484    484       G -> S (IN REF. 2).
SQ   SEQUENCE   563 AA;  60979 MW;  7764DB57 CRC32;

  Query Match           12.7%;  Score 437;  DB 2;  Length 563;
  Best Local Similarity  28.8%;  Pred. No. 2.88e-70;
  Matches    93;  Conservative   84;  Mismatches 123;  Indels  23;  Gaps  19;

Db    231 pfaalqkaqkgvtawddpykgnfgqlmalkqahpdlkilpsiggwtlsdpfff-mgdkvk 289
          :||:: : | : | |:|     : :: :||  :| || | :|||| ::   |  |     :
Qy     57 AFAGMTNHQLSTTEWNDE-T-LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATAN 114

Db    290 -rdrfvgsvkeflqtwkffdgvdidwefpggkganpnlgspqdgetyvllmkelramldq 348
           |: || |   ||:   | ||:|:|||:||: |: | :    | | :  |: :|   ::|
Qy    115 NRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV----DKERFTTLVQDLANAFQQ 168

Db    349 lsaetgr-kyeltsaisagkdkidkvaynvaq--nsmdhiflmsydfygpfdlknlghqt 405
           :  :|: :  |::|::||   :|  :|:|     ::| : ||:||| |: : |  || :
Qy    169 EAQTSGKERLLLSAAVPAGQTYVD-AGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNS 226

Db    406 alnapawkpdtayt-tvng-vnallaqgvkpgkvvvgtamygrgwtgvngyqnnipftgt 463
          :|      : :| :  |:: |   |  |  ::|:::| : |||: |  :: :  :   :|
Qy    227 PLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPAT 286

Db    464 atg-pvkgtwkngivdyrqiagqfmsgewqytydataeapyvfkpstgdlitfddarsvq 522
          ::| |   |   |:: | :: :  :::  |   :     ||:|:    : : |||  |
Qy    287 GSGTPGPFTKEGGMLAYYEVCS-WKGATKQ-RIQD-QKVPYIFR-DN-QWVGFDDVESFK 341

Db    523 akgkyvldkqlgglfsweidadn 545
          :|  |: :| |||   | :| |:
Qy    342 TKVSYLKQKGLGGAMVWALDLDD 364


RESULT    9
ID   CHIT_NPVAC     STANDARD;      PRT;   551 AA.
AC   P41684;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   PROBABLE ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
OS   AUTOGRAPHA CALIFORNICA NUCLEAR POLYHEDROSIS VIRUS (ACMNPV).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; BACULOVIRIDAE; EUBACULOVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C6;
RX   MEDLINE; 94303173.
RA   AYRES M.D., HOWARD S.C., KUZIO J., LOPEZ-FERBER M., POSSEE R.D.;
RL   VIROLOGY 202:586-605(1994).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; L22858; G559195; -.
DR   PROSITE; PS00014; ER_TARGET.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    551       PROBABLE ENDOCHITINASE.
FT   ACT_SITE    305    305       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    173    173       POTENTIAL.
FT   CARBOHYD    444    444       POTENTIAL.
FT   SITE        548    551       PREVENT SECRETION FROM ER (POTENTIAL).
SQ   SEQUENCE   551 AA;  61368 MW;  BDB124DB CRC32;

  Query Match           11.5%;  Score 395;  DB 2;  Length 551;
  Best Local Similarity  29.2%;  Pred. No. 1.25e-60;
  Matches    95;  Conservative   80;  Mismatches 122;  Indels  28;  Gaps  27;

Db    221 pwaavqkpqkgvsawn-ep-ykgnfgqlmaaklanphlkilpsiggwtlsdp-fyfmhdv 277
          : |:: : | : : || |  |  :|  |:  |: || || | :|||| ::   |  |
Qy     57 AFAGMTNHQLSTTEWNDETLYQ-EFNGLK--KM-NPKLKTLLAIGGWNFGTQKFTDMVAT 112

Db    278 -ekrnvfvdsvkeflqvwkffdgvdidwefpggkganps1gdadgdaktyillleelram 336
           ::|  ||:|   ||:   | ||:|:|||:||: |: |::  |   : : :  |:::|
Qy    113 ANNRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV-D-K-E-R-FTTLVQDLANA 165

Db    337 lddleaqt-grv-yeltsaisagydkiavvnya-e-aqkslgkiflmsydfkgawsntdl 392
          ::: |||| |:    |::|::|| :      |  :    :|  : ||:||| |:| :
Qy    166 FQQ-EAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVT- 222

Db    393 gyqttvya-ps-wnseelytthyavdallkqgvdpnkiivgvamygrgwtgvtnytndny 450
          |  : :|       :         ||:  |  |  ::|:|:|:: |||: |  ::
Qy    223 GHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVG 282

Db    451 fsgtgng-pgsgtwedgvvdyrqiqkdlnnyvytfdsaaqasyvfdkskgdlisfdsvds 509
          ::||:| ||: | | |:: | ::          ::     :|:| : : : ::|| |:|
Qy    283 APATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQ-KVPYIF-RDN-QWVGFDDVES 339

Db    510 vlgkvkyvdrnklgglfaweidadn 534
            || |: :: |||   | :| |:
Qy    340 FKTKVSYLKQKGLGGAMVWALDLDD 364


RESULT   10
ID   CHIA_ALTSO     STANDARD;      PRT;   820 AA.
AC   P32823;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE A PRECURSOR (EC 3.2.1.14) (CHI-A).
GN   CHIA.
OS   ALTEROMONAS SP. (STRAIN O-7).
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   VIBRONACEAE.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93106952.
RA   TSUJIBO H., ORIKOSHI H., TANNO H., FUJIMOTO K., MIYAMOTO K.,
RA   IMADA C., OKAMI Y., INAMORI Y.;
RL   J. BACTERIOL. 175:176-181(1993).
RN   [2]
RP   SEQUENCE OF 22-48, AND CHARACTERIZATION.
RX   MEDLINE; 93099552.
RA   TSUJIBO H., YOSHIDA Y., MIYAMOTO K., IMADA C., OKAMI Y., INAMORI Y.;
RL   CAN. J. MICROBIOL. 38:891-897(1992).
RN   [3]
RP   MUTAGENESIS.
RX   MEDLINE; 93379363.
RA   TSUJIBO H., ORIKOSHI H., IMADA C., OKAMI Y., MIYAMOTO K.,
RA   INAMORI Y.;
RL   BIOSCI. BIOTECHNOL. BIOCHEM. 57:1396-1397(1993).
CC   -!- THE OPTIMUM PH AND TEMPERATURE OF CHI-A ARE 8.0 AND 50 DEGREES
CC       CELSIUS, RESPECTIVELY.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- ENZYME REGULATION: STIMULATED BY MG2+; INHIBITED BY
CC       N-BROMOSUCCINIMIDE AND 2-HYDROXY-5-NITROBENZYL BROMIDE.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; D13762; G216207; -.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL.
FT   SIGNAL        1     21
FT   CHAIN        22    820       CHITINASE A.
FT   ACT_SITE    313    313       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   820 AA;  87346 MW;  838F6EE8 CRC32;

  Query Match           10.0%;  Score 345;  DB 2;  Length 820;
  Best Local Similarity  35.0%;  Pred. No. 2.40e-49;
  Matches    57;  Conservative   44;  Mismatches  47;  Indels  15;  Gaps  13;

Db    250 vysqlmalkqrypdlkilpsvggwtlsdp-fhgftnkan-rdtfvasvkqflktwkfydg 307
          :| :: :||   | || | ::||| ::   |      || |:||| |  :||:   | ||
Qy     76 LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSF-DG 134

Db    308 vdidwefpggdgpnpdlgdpindgpayvalmqelramldeleaet-grq-yeltsaigag 365
          :|:|||:||::|: | : |  ::   : :|:|:|   ::: ||:| |::   |::|: ||
Qy    135 LDLDWEYPGSQGS-PAV-D--KE-R-FTTLVQDLANAFQQ-EAQTSGKERLLLSAAVPAG 187

Db    366 ydkiedvdyqaaq--qymdyifamtydfygawnnetghqtgiy 406
            :   |  |:     | :|::  |:||| |:|:: ||| : :|
Qy    188 -QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLY 229


RESULT   11
ID   CHIB_SERMA     STANDARD;      PRT;   499 AA.
AC   P11797;
DT   01-OCT-1989 (REL. 12, CREATED)
DT   01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE B PRECURSOR (EC 3.2.1.14).
GN   CHIB.
OS   SERRATIA MARCESCENS.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   ENTEROBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 990 / QMB1466;
RX   MEDLINE; 89345110.
RA   HARPSTER M.H., DUNSMUIR P.;
RL   NUCLEIC ACIDS RES. 17:5395-5395(1989).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; X15208; G47228; -.
DR   PIR; S04856; S04856.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL.
FT   SIGNAL        1     41
FT   CHAIN        42    499       CHITINASE B.
FT   ACT_SITE    144    144       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   499 AA;  55464 MW;  9A585FFC CRC32;

  Query Match           9.8%;  Score 337;  DB 2;  Length 499;
  Best Local Similarity  34.0%;  Pred. No. 1.46e-47;
  Matches    54;  Conservative   47;  Mismatches  44;  Indels  14;  Gaps  10;

Db     78 ltalkahnpslrimfsiggwyysndlgvshanyvnavktpaartkfaqscvrimkdygfd 137
          : :||  || |: :::|||||   | :| :: :: : | |:  |  |  | :|::: |:||
Qy     80 FNGLKKMNPKLKTLLAIGGW---N-FG-TQ-KFTDMVATANNRQTFVNSAIRFLRKYSFD 133

Db    138 gvdidweyp--qaa-evd-g-fiaalqeirtllnqqtiadgrqalpyqltiagaggaffl 192
          |:|:|||||  |::  ||   | : :|::   : |:: : |:: |   |: | ::|  ::
Qy    134 GLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERL--LLSAAVPAGQTYV 191

Db    193 sryysklaqivapldyinlmtydlagpwekitnhqaalf 231
              :  |   ||::|||:||: |:|||:| | ::|:
Qy    192 DAGYE-VDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLY 229


RESULT   12
ID   CHIT_STRLI     STANDARD;      PRT;   619 AA.
AC   P36909;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE C PRECURSOR (EC 3.2.1.14).
GN   CHIC.
OS   STREPTOMYCES LIVIDANS.
OC   PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93294525.
RA   FUJII T., MIYASHITA K.;
RL   J. GEN. MICROBIOL. 139:677-686(1993).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- INDUCTION: BY CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS A BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC       (CBD).
DR   EMBL; D12647; G391910; -.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS00018; EF_HAND.
DR   PROSITE; PS00561; CBD_BACTERIAL.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    619       CHITINASE C.
FT   DOMAIN       35    140       CELLULOSE-BINDING.
FT   DOMAIN      148    230       FIBRONECTIN TYPE-III.
FT   DOMAIN      240    619       CATALYTIC.
FT   ACT_SITE    382    382       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   619 AA;  65200 MW;  23E3E1EC CRC32;

  Query Match           8.7%;  Score 301;  DB 2;  Length 619;
  Best Local Similarity  28.2%;  Pred. No. 1.25e-39;
  Matches    86;  Conservative   65;  Mismatches 129;  Indels  25;  Gaps  18;

Db    311 w-dqplrgnfnqlrklkakypnikilysfggwtwsgg-fpdavknpaafakschdlvedp 368
          | |: |  :|| |:|::   |::| | ::|||  :   | | |  :       :  :
Qy     71 WNDETLYQEFNGLKKMN---PKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFL 127

Db    369 rwadvfdgidldweypnacglscdetsapnafssmmkamraefgqdylita-a---vtad 424
          |    |||:|||||||| : | |   :   : |::::  :   | |:   ::     ::|
Qy    128 RKYS-FDGLDLDWEYPGSQG-S--PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAA 183
```

```
Db    425 gsdggk-idaa-dygeaskyidwynvmtydffgawak-ngptaphspltaydgipqqgfn 481
          : |   :||: :    :  :|  |:|:||| |:| |  |  :|        | :  ::|
Qy    184 VPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA-SLN 242

Db    482 tadamakfkskgvpadklligigfygrgwt-gvtqsap-ggtatgpat-gty--eagied 536
            |:     || || ||::|:  |||: | : :  :  |: |||::| | :  |:|:
Qy    243 VDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLA 302

Db    537 y-kvlknscpatgtiag--tayahcgsnwwsydtpatikskmdwaeqqglggaffwefsg 593
          | |     ::  |    :|    : | ::|   ::|:|:    | |||||  | :
Qy    303 YYEVCSWKGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDL 362

Db    594 dtang 598
          |   |
Qy    363 DDFAG 367


RESULT   13
ID   CHIT_STRPL     STANDARD;      PRT;   610 AA.
AC   P11220;
DT   01-JUL-1989 (REL. 11, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE 63 PRECURSOR (EC 3.2.1.14).
GN   CHTA.
OS   STREPTOMYCES PLICATUS.
OC   PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92192480.
RA   ROBBINS P.W., OVERBYE K., ALBRIGHT C., BENFIELD B., PERO J.;
RL   GENE 111:69-76(1992).
RN   [2]
RP   SEQUENCE OF 1-45 FROM N.A., AND SEQUENCE OF 31-45.
RX   MEDLINE; 88087127.
RA   ROBBINS P.W., ALBRIGHT C., BENFIELD B.;
RL   J. BIOL. CHEM. 263:443-447(1988).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- INDUCTION: BY CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS A BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC       (CBD).
DR   EMBL; M82804; G153216; -.
DR   EMBL; M18397; G153209; -.
DR   PIR; A29912; A29912.
DR   PIR; JH0573; JH0573.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS00018; EF_HAND.
DR   PROSITE; PS00561; CBD_BACTERIAL.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL.
FT   SIGNAL       1     30
FT   CHAIN       31    610       CHITINASE 63.
FT   DOMAIN      35    140       CELLULOSE-BINDING.
FT   DOMAIN     148    229       FIBRONECTIN TYPE-III.
FT   DOMAIN     236    610       CATALYTIC.
FT   ACT_SITE   383    383       PROTON DONOR (BY SIMILARITY).
FT   CONFLICT     3      3       F -> I (IN REF. 2).
SQ   SEQUENCE   610 AA;  63974 MW;  2F5E8E35 CRC32;

  Query Match            8.3%;  Score 287;  DB 2;  Length 610;
  Best Local Similarity  26.4%;  Pred. No. 1.37e-36;
  Matches    78;  Conservative   63;  Mismatches 134;  Indels  21;  Gaps  16;

Db    320 fnqlrnlkaeyphikilysfggwtwsgg-fpdavknpaafakschdlvedprwadvfdgi 378
          : ::  ||   | :| | ::|||  :   | | |  :       :  :   |    |||:
Qy     77 YQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYS-FDGL 135

Db    379 dldweypnacglscdetsapnafssmmkamraefgqdylita-a---vtadgsdggk-id 433
          ||||||| : | |   :   : |:::: :   | |:   ::     ::|  : |   :|
Qy    136 DLDWEYPGSQG-S--PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVD 192

Db    434 aa-dygeaskyidwynvmtydffgawak-ngptaphsplnaydgipqqgfttadamakfk 491
          |: :    : :|  |:|:||| |:| |  |  :|       | :  ::    |:
Qy    193 AGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA-SLNVDAAVQQWL 251

Db    492 skgvpadklligigfygrgwt-gvtqsap-ggtatgpaa-gty--eagiedy-kvlknsc 545
          || || ||::|:  |||: | : :  :  |: |||:: | :  |:|:  |  |
Qy    252 QKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKG 311

Db    546 patgtvag--tayahcgtnwwsydtpatikskmdwaeqqglggaffwefsgdttng 599
          ::   :     :|      | ::|   ::|:|:    | |||||  | :  |   |
Qy    312 ATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAG 367


RESULT   14
ID   KTXA_KLULA     STANDARD;      PRT;  1146 AA.
AC   P09805;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   KILLER TOXIN ALPHA AND BETA SUBUNITS PRECURSOR (RF2 PROTEIN)
DE   (ENDOCHITINASE (EC 3.2.1.14)).
OS   KLUYVEROMYCES LACTIS (YEAST).
OG   PLASMID PGKL-1.
OC   EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS 2359/152;
RA   SOR F., FUKUHARA H.;
RL   CURR. GENET. 9:147-155(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84297209.
RA   STARK M.J.R., MILEHAM A.J., ROMANOS M.A., BOYD A.;
RL   NUCLEIC ACIDS RES. 12:6011-6030(1984).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85037931.
RA   TOKUNAGA M., KAWAMURA A., HISHINUMA F.;
RL   NUCLEIC ACIDS RES. 12:7581-7597(1984).
RN   [4]
RP   IDENTIFICATION OF PROTEIN, AND SEQUENCE OF 30-44 AND 895-916.
RX   MEDLINE; 87004569.
RA   STARK M.J.R., BOYD A.;
RL   EMBO J. 5:1995-2002(1986).
RN   [5]
RP   SIMILARITY TO CHITINASE OF ALPHA-SUBUNIT.
RX   MEDLINE; 90259069.
RA   BRADSHAW H.D. JR.;
RL   NATURE 345:299-299(1990).
RN   [6]
RP   CHITINASE ACTIVITY OF ALPHA-SUBUNIT.
RX   MEDLINE; 91301161.
RA   BUTLER A.R., O'DONNEL R.W., MARTIN V.J., GOODAY G.W., STARK M.J.R.;
RL   EUR. J. BIOCHEM. 199:483-488(1991).
CC   -!- FUNCTION: THE ALPHA SUBUNIT IS A POTENT EXOCHITINASE. ALONG WITH
CC       THE BETA SUBUNIT IT PLAYS A ROLE IN THE INITIAL INTERACTION OF THE
CC       TOXIN WITH SENSITIVE CELLS AND ALLOW THE GAMMA SUBUNIT (THE ACTIVE
CC       TOXIN) TO GAIN ENTRY INTO THE CELL.
CC   -!- PTM: RF2 IS POTENTIALLY SPLIT BY MEMBRANE-BOUND BASIC AMINO ACID-
CC       SPECIFIC PEPTIDASE TO YIELD THE ALPHA AND BETA SUBUNITS.
CC   -!- SUBUNIT: THE KILLER TOXIN IS COMPOSED OF THREE SUBUNITS: ALPHA,
CC       BETA AND GAMMA.
CC   -!- SIMILARITY: LOCAL, TO OTHER CHITIN-BINDING PROTEINS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
DR   EMBL; X07127; G2829; -.
DR   EMBL; X00762; G2844; -.
DR   EMBL; X01095; G2849; -.
DR   PIR; S07915; S07915.
DR   HSSP; P02877; 1HEV.
DR   PROSITE; PS00026; CHITIN_BINDING.
```

```
DR   PROSITE; PS01095; CHITINASE_18.
KW   PLASMID; TOXIN; GLYCOPROTEIN; SIGNAL; HYDROLASE; GLYCOSIDASE;
KW   CHITIN DEGRADATION.
FT   SIGNAL        1     17       POTENTIAL.
FT   PROPEP       18     29       POTENTIAL.
FT   CHAIN        30    892       ALPHA SUBUNIT (PROBABLE).
FT   CHAIN       895   1146       BETA SUBUNIT (PROBABLE).
FT   DOMAIN      310    363       CHITIN-BINDING (BY SIMILARITY).
FT   DOMAIN      364    635       CATALYTIC.
FT   ACT_SITE    495    495       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    771    771       POTENTIAL.
FT   CARBOHYD    858    858       POTENTIAL.
FT   CARBOHYD    868    868       POTENTIAL.
FT   CARBOHYD    876    876       POTENTIAL.
FT   CARBOHYD    995    995       POTENTIAL.
FT   CARBOHYD   1082   1082       POTENTIAL.
FT   CARBOHYD   1117   1117       POTENTIAL.
SQ   SEQUENCE   1146 AA;  128936 MW;  CDD1620F CRC32;

  Query Match             7.6%;  Score 263;  DB 5;  Length 1146;
  Best Local Similarity  28.4%;  Pred. No. 1.89e-31;
  Matches    60;  Conservative   58;  Mismatches  74;  Indels  19;  Gaps  14;

Db    425 ddsafsksaflkvtssk-k-ipsfggwdfstspstytifrnavktdqnrntfannlinfm 482
          |:: :     ||  ::| | : ::|||:|:|    :|   : | |  || || |: | |:
Qy     73 DETLYQEFNGLKKMNPKLKTLLAIGGWNFGT-Q-KFT---DMVATANNRQTFVNSAIRFL 127

Db    483 nkynldgidldweypgapdipdipaddsssgsnyltflkllkgkmps-gktl-siaipss 540
           ||::||:|||||||||:   | :  :  ::    |:      :  ::  : | | |:|::
Qy    128 RKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG 187

Db    541 ywylkn-fpisdiqntvdymvymtydihgiwe--ygkans-yi---nchtprk-eiedai 592
            |:   : :  |   :|::  |:||:|| ||   |  :: |    :  ::   ::: |:
Qy    188 QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAV 247

Db    593 km-ldkagvkfnkvfggvanygrsykmvntn 622
             |:| |   :|:: |:: ||||: : :::
Qy    248 QQWLQK-GTPASKLILGMPTYGRSFTLASSS 277


RESULT   15
ID   GP39_BOVIN     STANDARD;      PRT;    36 AA.
AC   P30922;
DT   01-JUL-1993 (REL. 26, CREATED)
DT   01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   CARTILAGE GLYCOPROTEIN-39 (GP-39) (39 KD WHEY PROTEIN) (FRAGMENT).
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; ARTIODACTYLA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 88106603.
RA   REJMAN J.J., HURLEY W.L.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 150:329-334(1988).
CC   -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE CAPACITY OF CELLS TO
CC       RESPOND TO AND COPE WITH CHANGES IN THEIR ENVIRONMENT.
CC   -!- PTM: GLYCOSYLATED.
CC   -!- TISSUE SPECIFICITY: MAMMARY SECRETIONS COLLECTED DURING THE
CC       NONLACTATING PERIOD.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
DR   PIR; A27682; A27682.
KW   GLYCOPROTEIN.
FT   NON_TER      36     36
SQ   SEQUENCE   36 AA;  4264 MW;  F3935740 CRC32;

  Query Match             5.0%;  Score 172;  DB 4;  Length 36;
  Best Local Similarity  48.6%;  Pred. No. 4.12e-13;
  Matches    17;  Conservative   11;  Mismatches   7;  Indels   0;  Gaps   0;

Db      2 klixyytswsqyregdgsxfpdaidpflxthviys 36
          ||: |:|:|:|||:|::  :|  :|| | ||:||:
Qy     23 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYA 57


Search completed: Fri Jun 27 16:48:55 1997
Job time : 128 secs.
```

```
*****************************************************************************
                     MPSRCH (TM)
*****************************************************************************

        Release 2.1D John F. Collins, Biocomputing Research Unit.
        Copyright (c) 1993, 1994, 1995  University of Edinburgh, U.K.
             Distribution rights by IntelliGenetics, Inc.

MPsrch_nn   n.a. - n.a. database search, using Smith-Waterman algorithm

Run on:          Sun Jun 29 20:19:14 1997;  MasPar time 1536.65 Seconds
                                            1198.286 Million cell updates/sec
Tabular output not generated.

Title:           >US-08-663-618A-3
Description:     (1-1656) from US08663618A.seq
Perfect Score:   1656
N.A. Sequence:       1 GCTGCAGCCTGCCGCTGAGC..........TGTGCCCCTCAAAAAAAAAA 1656
     Comp:               CGACGTCGGACGGCGACTCG..........ACACGGGGAGTTTTTTTTTT

Scoring table:   TABLE default
                 Gap 6

Nmatch    STD :  Dbase 0;  Query 0

Searched:        333249 seqs, 555961234 bases x 2

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database:        embl-new11
                   1:BCT 2:FUN 3:GEN 4:HUM1 5:HUM2 6:HUM3 7:INV1 8:INV2
                   9:INV3 10:INV4 11:INV5 12:INV6 13:INV7 14:ORG 15:MAM
                   16:VRT 17:PLN 18:PRO1 19:PRO2 20:ROD 21:SYN 22:UNC
                   23:VIR1 24:VIR2
Database:        genbank97
                   25:BCT1 26:BCT2 27:BCT3 28:BCT4 29:BCT5 30:BCT6 31:BCT7
                   32:BCT8 33:BCT9 34:GEN1 35:GEN2 36:HTG 37:INV1 38:INV2
                   39:INV3 40:INV4 41:INV5 42:INV6 43:INV7 44:INV8 45:INV9
                   46:MAM1 47:MAM2 48:MAM3 49:VRT1 50:VRT2 51:VRT3 52:PAT1
                   53:PAT2 54:PAT3 55:PAT4 56:PHG 57:PLN1 58:PLN2 59:PLN3
                   60:PLN4 61:PLN5 62:PLN6 63:PLN7 64:PLN8 65:PLN9 66:PLN10
                   67:PRI1 68:PRI2 69:PRI3 70:PRI4 71:PRI5 72:PRI6 73:PRI7
                   74:PRI8 75:PRI9 76:PRI10 77:PRI11 78:PRI12 79:PRI13
                   80:PRI14 81:ROD1 82:ROD2 83:ROD3 84:ROD4 85:ROD5 86:ROD6
                   87:ROD7 88:ROD8 89:STR 90:SYN 91:UNA 92:VRL1 93:VRL2
                   94:VRL3 95:VRL4 96:VRL5 97:VRL6 98:VRL7 99:VRL8 100:VRL9
Database:        genbank-new11
                   101:BCT 102:GEN 103:INV1 104:INV2 105:MAM 106:VRT
                   107:PHG 108:PLN 109:PRI1 110:PRI2 111:ROD 112:SYN
                   113:UNA 114:VRL
Database:        u-embl48_97
                   115:part1 116:part2

Statistics:      Mean 11.656;  Variance 4.533;  scale 2.572

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	1609	97.2	1633	74	HSU29615	Human chitotriosidase	0.00e+00
	2	182	11.0	1741	76	HUMHA3G	Human glycoprotein mR	1.14e-151
	3	157	9.5	1418	74	HSU49835	Human YKL-39 precurso	1.32e-125
	4	157	9.5	1434	74	HSU58514	Human chitinase precu	1.32e-125
	5	155	9.4	1500	74	HSU58515	Human chitinase (HUMT	1.56e-123
	6	154	9.3	1733	48	SSGP38KD	S.scrofa 38kDa hepari	1.69e-122
	7	154	9.3	1733	48	SSU19900	Sus scrofa 38kDa hepa	1.69e-122
	8	154	9.3	1733	15	SSGP38KD	S.scrofa 38kDa hepari	1.69e-122
	9	133	8.0	1616	81	MMBRP39	M.musculus mRNA for B	6.95e-101
	10	121	7.3	1506	20	MMD757	House mouse; Musculus	1.14e-88
	11	121	7.3	1506	81	D87757	House mouse; Musculus	1.14e-88
	12	103	6.2	2022	48	SSU43490	Sus scrofa oviductal	1.38e-70
	13	95	5.7	2034	47	OAU16719	Ovis aries estrogen d	1.15e-62
	14	91	5.5	1596	47	OAU17988	Ovis aries 92kD oestr	9.79e-59
	15	91	5.5	2228	67	BABEDOSG	Papio hamadryas anubi	9.79e-59
	16	91	5.5	2228	15	PAEDOSG	Papio hamadryas anubi	9.79e-59
	17	91	5.5	2525	20	MMOSG	Mouse mRNA for oviduc	9.79e-59
	18	91	5.5	2525	85	MUSOGP	Mouse mRNA for oviduc	9.79e-59
	19	87	5.3	2198	73	HSU09550	Human oviductal glyco	7.85e-55
	20	83	5.0	1994	46	BOV95BOGP	Bovine mRNA for 95 kD	5.93e-51
	21	67	4.0	2387	20	MAOSG	Hamster mRNA for ovid	9.19e-36
	22	67	4.0	2387	81	HAMOGP	Hamster mRNA for ovid	9.19e-36
	23	59	3.6	2276	81	MAU15048	Mesocricetus auratus	2.04e-28
	24	50	3.0	2452	43	MSU02270	Manduca sexta chitina	1.98e-20
	25	45	2.7	443	72	HSOGP03	Human oviductal glyco	3.70e-16
	26	45	2.7	994	83	MMU56900	Mus musculus chitinas	3.70e-16
	27	41	2.5	506	72	HSOGP07	Human oviductal glyco	7.55e-13
	28	42	2.5	3060	29	KURCHISHI	Kurthia zopfii gene f	1.15e-13
	29	40	2.4	2401	25	BACCHIA3	B.circulans chitinase	4.88e-12
	30	37	2.2	2078	44	PEUPJCHI1	Penaeus japonicus adu	1.18e-09
c	31	36	2.2	25083	7	CEC04F6	Caenorhabditis elegan	7.05e-09
c	32	36	2.2	25083	38	CELC04F6	Caenorhabditis elegan	7.05e-09
	33	34	2.1	1322	7	BPU59689	Brugia pahangi microf	2.36e-07
	34	34	2.1	1322	37	BPU59689	Brugia pahangi microf	2.36e-07
	35	34	2.1	1322	103	BPU59689	Brugia pahangi microf	2.36e-07
	36	34	2.1	1654	37	BRPCHIT	Brugia malayi chitina	2.36e-07
	37	33	2.0	1836	32	SHCHITB	S.marcescens (BJL200)	1.32e-06
	38	31	1.9	1320	54	I14183	Sequence 14 from pate	3.82e-05
	39	32	1.9	1364	37	BPU59690	Brugia pahangi microf	7.20e-06
	40	32	1.9	1364	7	BPU59690	Brugia pahangi microf	7.20e-06
	41	32	1.9	1364	103	BPU59690	Brugia pahangi microf	7.20e-06
	42	31	1.9	1364	54	I14182	Sequence 13 from pate	3.82e-05
	43	31	1.9	1405	54	I14180	Sequence 9 from paten	3.82e-05
	44	32	1.9	1700	32	SMCHIB	Serratia marcescens c	7.20e-06
	45	31	1.9	1904	57	AACHI1A	A.album chi1 gene for	3.82e-05

ALIGNMENTS

```
RESULT     1
LOCUS       HSU29615      1633 bp    mRNA            PRI       04-NOV-1995
DEFINITION  Human chitotriosidase precursor mRNA, complete cds.
ACCESSION   U29615
NID         g1050957
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1633)
  AUTHORS   Boot,R.G., Renkema,G.H., Strijland,A., van Zonneveld,A.J. and
            Aerts,J.M.
  TITLE     Cloning of a cDNA encoding chitotriosidase, a human chitinase
            produced by macrophages
  JOURNAL   J. Biol. Chem. 270 (44), 26252-26256 (1995)
  MEDLINE   96064695
REFERENCE   2  (bases 1 to 1633)
  AUTHORS   Boot,R.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-1995) Rolf G. Boot, Department of Biochemistry,
            AMC, E.C. Slater, Institute, University of Amsterdam, Meibergdreef
            15, 1105 AZ, Amsterdam, The Netherlands
```

```
FEATURES             Location/Qualifiers
     source          1..1633
                     /organism="Homo sapiens"
                     /clone="Ch1.1"
                     /clone_lib="macrophage-1 cDNA library of Rolf G. Boot"
                     /cell_type="macrophage"
     sig_peptide     13..75
     CDS             13..1413
                     /codon_start=1
                     /product="chitotriosidase precursor"
                     /db_xref="PID:g1050958"
                     /translation="MVRSVAWAGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPK
                     DLDPSLCTHLIYAFAGMTNHQLSTTEWNDETLYQEFNGLKKMNPKLKTLLAIGGWNFG
                     TQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQ
                     DLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSW
                     EKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASS
                     SDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQKVPYIFRDNQWVGF
                     DDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYPLIQTLRQELSLPYLPS
                     GTPELEVPKPGQPSEPEHGPSPGQDTFCQGKADGLYPNPRERSSFYSCAAGRLFQQSC
                     PTGLVFSNSCKCCTWN"
     mat_peptide     76..1410
                     /note="human chitinase"
                     /product="chitotriosidase"
     polyA_signal    1605..1610
     polyA_site      1633
                     /note="10 A nucleotides"
BASE COUNT      354 a    489 c    443 g    347 t
ORIGIN

  Query Match           97.2%;  Score 1609;  DB 74;  Length 1633;
  Best Local Similarity  99.8%;  Pred. No. 0.00e+00;
  Matches  1628;  Conservative    0;  Mismatches   1;  Indels   3;  Gaps   3;

Db       1 ctgagctgcatcatggtgcggtctgtggcctgggcaggtttcatggtcctgctgatgatc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      15 CTGAGCTGCATCATGGTGCGGTCTGTGGCCTGGGCAGGTTTCATGGTCCTGCTGATGATC 74

Db      61 ccatggggctctgctgcaaaactggtctgctacttcaccaactgggcccagtacagacag 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      75 CCATGGGGCTCTGCTGCAAAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAG 134

Db     121 ggggaggctcgcttcctgcccaaggacttggaccccagcctttgcacccacctcatctac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     135 GGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTAC 194

Db     181 gccttcgctggcatgaccaaccaccagctgagcaccactgagtggaatgacgagactctc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     195 GCCTTCGCTGGCATGACCAACCACCAGCTGAGCACCACTGAGTGGAATGACGAGACTCTC 254

Db     241 taccaggagttcaatggcctgaagaagatgaatcccaagctgaagaccctgttagccatc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     255 TACCAGGAGTTCAATGGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATC 314

Db     301 ggaggctggaatttcggcactcagaagttcacagatatggtagccacggccaacaaccgt 360
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Qy     315 GGAGGCTGGAATTTCAGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGT 374

Db     361 cagacctttgtcaactcggccatcaggtttctgcgcaaatacagctttgacggccttgac 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     375 CAGACCTTTGTCAACTCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGAC 434

Db     421 cttgactgggagtacccaggaagccaggggagccctgccgtagacaaggagcgcttcaca 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     435 CTTGACTGGGAGTACCCAGGAAGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACA 494

Db     481 accctggtacaggacttggccaatgccttccagcaggaagcccagacctcagggaaggaa 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     495 ACCCTGGTACAGGACTTGGCCAATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAA 554

Db     541 cgccttcttctgagtgcagcggttccagctgggcagacctatgtggatgctggatacgag 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     555 CGCCTTCTTCTGAGTGCAGCGGTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAG 614

Db     601 gtggacaaaatcgcccagaacctggattttgtcaaccttatggcctacgacttccatggc 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     615 GTGGACAAAATCGCCCAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGC 674

Db     661 tcttgggagaaggtcacgggacataacagccccctctacaagaggcaagaagagagtggt 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     675 TCTTGGGAGAAGGTCACGGGACATAACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGT 734

Db     721 gcagcagccagcctcaacgtggatgctgctgtgcaacagtggctgcagaaggggacccct 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     735 GCAGCAGCCAGCCTCAACGTGGATGCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCT 794

Db     781 gccagcaagctgatccttggcatgcctacctacggacgctccttcacactggcctcctca 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     795 GCCAGCAAGCTGATCCTTGGCATGCCTACCTACGGACGCTCCTTCACACTGGCCTCCTCA 854

Db     841 tcagacaccagagtgggggccccagccacagggtctggcactccaggccccttcaccaag 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     855 TCAGACACCAGAGTGGGGGCCCCAGCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAG 914

Db     901 gaaggagggatgctggcctactatgaagtctgctcctggaaggggccaccaaacagaga 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     915 GAAGGAGGGATGCTGGCCTACTATGAAGTCTGCTCCTGGAAGGGGCCACCAAACAGAGA 974

Db     961 atccaggatcagaaggtgccctacatcttccgggacaaccagtgggtgggctttgatgat 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     975 ATCCAGGATCAGAAGGTGCCCTACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGAT 1034

Db    1021 gtggagagcttcaaaaccaaggtcagctatctgaagcagaagggactgggcggggccatg 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1035 GTGGAGAGCTTCAAAACCAAGGTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATG 1094

Db    1081 gtctgggcactggacttagatgactttgccggcttctcctgcaaccagggccgataccc 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1095 GTCTGGGCACTGGACTTAGATGACTTTGCCGGCTTCTCCTGCAACCAGGGCCGATACCCC 1154

Db    1141 ctcatccagacgctacggcaggaactgagtcttccatacttgccttcaggcaccccagag 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1155 CTCATCCAGACGCTACGGCAGGAACTGAGTCTTCCATACTTGCCTTCAGGCACCCCAGAG 1214

Db    1201 cttgaagttccaaaaccaggtcagccctctgaacctgagcatggccccagccctggacaa 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1215 CTTGAAGTTCCAAAACCAGGTCAGCCCTCTGAACCTGAGCATGGCCCCAGCCCTGGACAA 1274

Db    1261 gacacgttctgccagggcaaagctgatgggctctatcccaatcctcgggaacggtccagc 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1275 GACACGTTCTGCCAGGGCAAAGCTGATGGGCTCTATCCCAATCCTCGGGAACGGTCCAGC 1334

Db    1321 ttctacagctgtgcagcggggcggctgttccagcaaagctgcccgacaggcctggtgttc 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1335 TTCTACAGCTGTGCAGCGGGGCGGCTGTTCCAGCAAAGCTGCCCGACAGGCCTGGTGTTC 1394

Db    1381 agcaactcctgcaaatgctgcacctggaattgagtcg-taaagcccctccagtcc-agct 1438
           |||||||||||||||||||||||||||||||||||| |||||||||||||||| |||| 
Qy    1395 AGCAACTCCTGCAAATGCTGCACCTGGAATTGAGTCGCTAAAGCCCCTCCAGTCCCAGCT 1454

Db    1439 ttgaggctgggcccaggatcactctacagcctgcctcctgggttttcc-tgggggccgca 1497
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||| 
Qy    1455 TTGAGGCTGGGCCCAGGATCACTCTACAGCCTGCCTCCTGGGTTTTCCCTGGGGGCCGCA 1514

Db    1498 atctggctcctgcaggcctttctgtggtcttcctttatccaggctttctgctctcagcct 1557
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1515 ATCTGGCTCCTGCAGGCCTTTCTGTGGTCTTCCTTTATCCAGGCTTTCTGCTCTCAGCCT 1574

Db    1558 tgccttccttttttctgggtctcctgggctgccccttcacttgcaaaataaatctttgg 1617
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1575 TGCCTTCCTTTTTTCTGGGTCTCCTGGGCTGCCCCTTTCACTTGCAAAATAAATCTTTGG 1634

Db    1618 tttgtgcccctc 1629
           ||||||||||||
Qy    1635 TTTGTGCCCCTC 1646
```

```
RESULT    2
LOCUS       HUMHA3G      1741 bp    mRNA            PRI       24-JAN-1994
DEFINITION  Human glycoprotein mRNA, complete cds.
ACCESSION   M80927
NID         g348911
KEYWORDS    glycoprotein.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1741)
  AUTHORS   Hakala,B.E., White,C. and Recklies,A.D.
  TITLE     Human cartilage gp-39, a major secretory product of articular
            chondrocytes and synovial cells, is a mammalian member of a
            chitinase protein family
  JOURNAL   J. Biol. Chem. 268 (34), 25803-25810 (1993)
  MEDLINE   94064658
FEATURES             Location/Qualifiers
     source          1..1741
                     /organism="Homo sapiens"
                     /cell_type="chondrocyte"
                     /tissue_type="cartilage"
     5'UTR           1..71
     sig_peptide     72..134
                     /codon_start=1
     CDS             72..1223
                     /note="articular 39kDa glycoprotein"
                     /codon_start=1
                     /evidence=experimental
                     /product="glycoprotein"
                     /db_xref="PID:g348912"
                     /translation="MGVKASQTGFVVLVLLQCCSAYKLVCYYTSWSQYREGDGSCFPD
                     ALDRFLCTHIIYSFANISNDHIDTWEWNDVTLYGMLNTLKNRNPNLKTLLSVGGWNFG
                     SQRFSKIASNTQSRRTFIKSVPPFLRTHGFDGLDLAWLYPGRRDKQHFTTLIKEMKAE
                     FIKEAQPGKKQLLLSAALSAGKVTIDSSYDIAKISQHLDFISIMTYDFHGAWRGTTGH
                     HSPLFRGQEDASPDRFSNTDYAVGYMLRLGAPASKLVMGIPTFGRSFTLASSETGVGA
                     PISGPGIPGRFTKEAGTLAYYEICDFLRGATVHRTLGQQVPYATKGNQWVGYDDQESV
                     KSKVQYLKDRQLAGAMVWALDLDDFQGSFCGQDLRFPLTNAIKDALAAT"
     mat_peptide     135..1220
                     /codon_start=1
                     /product="glycoprotein"
     3'UTR           1224..1741
     polyA_signal    1715..1720
     polyA_site      1741
BASE COUNT      416 a    510 c    447 g    368 t
ORIGIN
```

Query Match 11.0%; Score 182; DB 76; Length 1741;
Best Local Similarity 63.8%; Pred. No. 1.14e-151;
Matches 700; Conservative 0; Mismatches 374; Indels 24; Gaps 10;

```
Db      94 caggctttgtggtcctggtgctgctccagtgctgctctgcatacaaactggtctgctact 153
Qy      49 CAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTGCTACT 108

Db     154 acaccagctggtcccagtaccgggaaggcgatgggagctgcttcccagatgcccttgacc 213
Qy     109 TCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACC 168

Db     214 gcttcctctgtacccacatcatctacagctttgccaatataagcaacgatcacatcgaca 273
Qy     169 CCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCTGAGCA 228

Db     274 cctgggagtggaatgatgtgacgctctacggcatgctcaacacactcaagaacaggaacc 333
Qy     229 CCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGATGAATC 288

Db     334 ccaacctgaagactctcttgtctgtcggaggatggaactttgggtctcaaagattttcca 393
Qy     289 CCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCAGCACTCAGAAGTTCACAG 348

Db     394 agatagcctccaacacccagagtcgccggactttcatcaagtcagtaccgccattcctgc 453
```

```
Qy     349 ATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTTTCTGC 408

Db     454 gcacccatggctttgatgggctggaccttgcctggctctaccctgga---c---ggag-- 505
Qy     409 GCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGCCAGGGGAGCC 468

Db     506 -------agacaaacagcattttaccaccctaatcaaggaaatgaaggccgaatttataa 558
Qy     469 CTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAATGCCTTCCAGC 528

Db     559 aggaagcccag--c-cagggaaaaagcagctcctgctcagcgcagcactgtctgcgggga 615
Qy     529 AGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCCAGCTGGGC 588

Db     616 aggtcaccattgacagcagctatgacattgccaagatatcccaacacctggatttcatta 675
Qy     589 AGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATTTTGTCA 648

Db     676 gcatcatgacctacgattttcatggagcctggcgtgggaccacaggccatcacagtcccc 735
Qy     649 ACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACATAACAGCCCCC 708

Db     736 tgttccgaggtcaggaggatgcaagtcctgacagattcagcaacactgactatgctgtgg 795
Qy     709 TCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGATGCTGCTGTGC 768

Db     796 ggtacatgttgaggctgggggctcctgccagtaagctggtgatgggcatccccaccttcg 855
Qy     769 AACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATGCCTACCTACG 828

Db     856 ggaggagcttcactctggcttcttc-tgagactg--gtgttggagccccaatctcaggac 912
Qy     829 GACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCAGCCACAGGGT 888

Db     913 cgggaattccaggccggttcaccaaggaggcagggacccttgcctactatgagatctgtg 972
Qy     889 CTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTATGAAGTCTGCT 948

Db     973 acttcctccgcggagccacagtccatagaaccctcggccagcaggtcccctatgccacca 1032
Qy     949 CCTGGAAG-G-GG-GCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCCTACATCTTCC 1005

Db    1033 agggcaaccagtgggtaggatacgacgaccaggaaagcgtcaaaagcaaggtgcagtacc 1092
Qy    1006 GGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAGGTCAGCTATC 1065

Db    1093 tgaaggataggcagctggcaggcgccatggtatgggccctggacctggatgacttccagg 1152
Qy    1066 TGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGATGACTTTGCCG 1125

Db    1153 gctccttctgcggccagg 1170
Qy    1126 GCTTCTCCTGCAACCAGG 1143
```

```
RESULT    3
LOCUS       HSU49835     1418 bp    mRNA            PRI       25-JUL-1996
DEFINITION  Human YKL-39 precursor mRNA, complete cds.
ACCESSION   U49835
NID         g1457940
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1418)
  AUTHORS   Hu,B., Trinh,K., Figueira,W.F. and Price,P.A.
  TITLE     Isolation and sequence of a novel human chondrocyte protein related
            to mammalian members of the chitinase protein family
  JOURNAL   J. Biol. Chem. 271 (32), 19415-19420 (1996)
  MEDLINE   96325055
REFERENCE   2  (bases 1 to 1418)
  AUTHORS   Price,P.A.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1996) Paul A. Price, Biology, 0322, University of
            California at San Diego, 9500 Gilman Drive, La Jolla, CA
            92093-0322, USA
FEATURES             Location/Qualifiers
     source          1..1418
                     /organism="Homo sapiens"
                     /tissue_type="articular cartilage"
     CDS             36..1193
                     /note="similar to human glycoprotein encoded by Genbank
                     Accession Number M80927, human chitotriosidase precursor
                     encoded by Genbank Accession Number U29615, human
                     oviductal glycoprotein encoded by Genbank Accession Number
                     U09550, and to mouse secretory protein YM-1 precursor
                     encoded by Genbank Accession Number S27879"
                     /codon_start=1
                     /product="YKL-39 precursor"
                     /db_xref="PID:g1457941"
                     /translation="MDQKSLWAGVVVLLLLQGGSAYKLVCYFTNWSQDRQEPGKFTPE
                     NIDPFLCSHLIYSFASIENNKVIIKDKSEVMLYQTINSLKTKNPKLKILLSIGGYLFG
                     SKGFHPMVDSSTSRLEFINSIILFLRNHNFDGLDVSWIYPDQKENTHFTVLIHELAEA
                     FQKDFTKSTKERLLLTAGVSAGRQMIDNSYQVEKLAKDLDFINLLSFDFHGSWEKPLI
                     TGHNSPLSKGWQDRGPSSYYNVEYAVGYWIHKGMPSEKVVMGIPTYGHSFTLASAETT
                     VGAPASGPGAAGPITESSGFLAYYEICQFLKGAKITRLQDQQVPYAVKGNQWVGYDDV
                     KSMETKVQFLKNLNLGGAMIWSIDMDDFTGKSCNQGPYPLVQAVKRSLGSL"
     sig_peptide     36..98
     mat_peptide     99..1190
                     /product="YKL-39"
     polyA_site      1418
                     /note="18 A nucleotides"
BASE COUNT      370 a    357 c    337 g    354 t
ORIGIN

  Query Match             9.5%;  Score 157;  DB 74;  Length 1418;
  Best Local Similarity  62.9%;  Pred. No. 1.32e-125;
  Matches   710;  Conservative    0;  Mismatches 391;  Indels  27;  Gaps  10;

Db      53 ctgggcaggtgtagtggtcttgctgcttctccagggaggatctgcctacaaactggtttg 112
           |||||||||| |  ||||| ||||| |  |||   | || |||||    |||||||| ||
Qy      44 CTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTG 103

Db     113 ctactttaccaactggtcccaggaccggcaggaaccaggaaaattcacccctgagaatat 172
           |||||| |||||||||| ||||| || | ||||     |     |||   ||  || |  |
Qy     104 CTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTT 163

Db     173 tgacccctttcctatgctctcatctcatctattcattcgccagcatcgaaaacaacaaggt 232
           ||||||  ||| ||| | || ||||||||  | |||||  ||||    ||| || || |
Qy     164 GGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCT 223

Db     233 tatcatcaaggacaagagtgaagtgatgctctaccagaccatcaacagtctcaaaaccaa 292
           | || ||  ||   || ||| | ||  |||||||||    ||||  | || || |  |
Qy     224 GAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGAT 283

Db     293 gaatcccaaactgaaaattctcttgtccattggagggtacctgtttggttccaaagggtt 352
           ||||||||| ||||| |  || ||  |||| ||||| |     ||  |  |  |   |||
Qy     284 GAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCAGCACTCAGAAGTT 343

Db     353 ccaccctatggtggattcttctacatcacgcttggaattcattaactccataatcctgtt 412
           |     ||||||  |   |  | |     ||  |  ||  | |||||    |||  |||
Qy     344 CACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTT 403

Db     413 tctgaggaaccataactttgatggactggatgtaagctggatctacccag--atc-agaa 469
           |||| | ||  | | |||||| || || ||  |   ||||   |||||||  | | ||
Qy     404 TCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGCCAGGG 463

Db     470 -agaa-------a-acactcat---ttcactgtgctgattcatgagttagcagaagcctt 517
            ||       | |||   |    |||||    ||| | || || || ||  | |||||
Qy     464 GAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAATGCCTT 523

Db     518 tcagaaggacttcacaaaatccaccaaggaaaggcttctcttgactgcgggcgtatctgc 577
           |||  ||||    |   |  ||    |||||| | |||||  ||| ||| |  ||  | ||
Qy     524 CCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCCAGC 583

Db     578 agggaggcaaatgattgataacagctatcaagttgagaaactggcaaaagatctggattt 637
           |||  |       | |||    | ||  | || || ||| | ||  |  | ||||||||
Qy     584 TGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATTT 643

Db     638 catcaacctcctgtcctttgacttccatgggtcttgggaaaagccccttatcactggcca 697
           |||||||  || |||  |||||||||||| |||||||| |||       |||| || ||
Qy     644 TGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGG------TCACGGGACA 697

Db     698 caacagccctctgagcaaggggtggcaggacagagggccaagctcctactacaatgtgga 757
           ||||||||  ||   |||| ||    | || || ||  ||    ||  |  ||| |||||
Qy     698 TAACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGA 757

Db     758 atatgctgtggggtactggatacataagggaatgccatcagagaaggtggtcatgggcat 817
           |||||||    | ||| | || ||||| |  ||  |    ||| || || | |||||
Qy     758 TGCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCAT 817

Db     818 ccccacatatgggcactccttcacactggcctc-tg--cagaaaccaccgtgggggcccc 874
           || || || || | |||||||||||||||||| |   |||| |||| ||||||||||||
Qy     818 GCCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCC 877

Db     875 tgcctctggccctggagctgctggacccatcacagagtcttcaggcttcctggcctatta 934
           |||  | ||  |||| || | || ||| |||| ||     |||  | |||||||| ||
Qy     878 AGCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTA 937

Db     935 tgagatctgccagttcctgaaaggagccaagatcacgcgcctccaggatcagcaggttcc 994
           |||  |||||   | |  ||| || ||||  |    | |  ||||||||||| |||| ||
Qy     938 TGAAGTCTGC---TCCTGGAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCC 994

Db     995 ctacgcagtcaaggggaaccagtgggtgggctatgatgatgtgaagagtatggagaccaa 1054
           ||||   ||  ||  |||||||||||||||| ||||||||| ||||  |  | |||||
Qy     995 CTACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAA 1054

Db    1055 ggttcagttcttaaagaatttaaacctgggaggagccatgatctggtctattgacatgga 1114
           |||    |   | ||| |       ||||| || |||||| ||||| |  | ||| | ||
Qy    1055 GGTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGA 1114

Db    1115 tgacttcactggcaaatcctgcaaccagggcccttaccctcttgtcca 1162
           ||||||  | |||  ||||||||||||||||  ||||| ||  ||||
Qy    1115 TGACTTTGCCGGCTTCTCCTGCAACCAGGGCCGATACCCCCTCATCCA 1162


RESULT    4
LOCUS       HSU58514     1434 bp    mRNA            PRI       24-JUL-1996
DEFINITION  Human chitinase precursor (HUMTCHIT) mRNA, exon 1a form, complete
            cds.
ACCESSION   U58514
NID         g1439565
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1434)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1434)
  AUTHORS   Grossman,A., Matsuyama,T., Baker,E., Waterhouse,P., Sutherland,G.R.
            and Mak,T.W.
  TITLE     Cloning of a novel lymphoid restricted human chitinase and
            localization to 1p13.3
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1434)
  AUTHORS   Grossman,A., Matsuyama,T., Baker,E., Waterhouse,P., Sutherland,G.R.
            and Mak,T.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1996) Medical Biophysics, Ontario Cancer
            Institute, Princess Margaret Hospital, 610 University Avenue,
            Toronto, Ontario M5G 2M9, Canada
```

```
FEATURES             Location/Qualifiers
     source          1..1434
                     /organism="Homo sapiens"
                     /map="1p13.3"
                     /chromosome="1"
                     /tissue_type="Soares placenta"
                     /dev_stage="8 to 9 weeks"
                     /clone="EST 257753"
     5'UTR           1..36
                     /gene="HUMTCHIT"
     exon            1..76
                     /gene="HUMTCHIT"
                     /note="exon 1a; first alternatively splice exon 1;
                     contains a Kozak consensus initiator methionine"
                     /number=1
     sig_peptide     37..114
                     /gene="HUMTCHIT"
     CDS             37..1209
                     /gene="HUMTCHIT"
                     /note="first alternatively spliced form, see GenBank
                     Accession Number U58515 for the second alternatively
                     spliced form"
                     /codon_start=1
                     /product="chitinase precursor"
                     /db_xref="PID:g1439566"
                     /translation="MGATTMDQKSLWAGVVVLLLLQGGSAYKLVCYFTNWSQDRQEPG
                     KFTPENIDPFLCSHLIYSFASIENNKVIIKDKSEVMLYQTINSLKTKNPKLKILLSIG
                     GYLFGSKGFHPMVDSSTSRLEFINSIILFLRNHNFDGLDVSWIYPDQKENTHFTVLIH
                     ELAEAFQKDFTKSTKERLLLTAGVSAGRQMIDNSYQVEKLAKDLDFINLLSFDFHGSW
                     EKPLITGHNSPLSKGWQDRGPSSYYNVEYAVGYWIHKGMPSEKVVMGIPTYGHSFTLA
                     SAETTVGAPASGPGAAGPITESSGFLAYYEICQFLKGAKITRLQDQQVPYAVKGNQWV
                     GYDDVKSMETKVQFLKNLNLGGAMIWSIDMDDFTGKSCNQGPYPLVQAVKRSLGSL"
     exon            77..106
                     /gene="HUMTCHIT"
                     /number=2
     exon            107..309
                     /gene="HUMTCHIT"
                     /number=3
     mat_peptide     115..1206
                     /gene="HUMTCHIT"
                     /note="first alternatively spliced form, see GenBank
                     Accession Number U58515 for the second alternatively
                     spliced form"
                     /product="chitinase"
     exon            310..365
                     /gene="HUMTCHIT"
                     /number=4
     exon            366..516
                     /gene="HUMTCHIT"
                     /number=5
     exon            517..643
                     /gene="HUMTCHIT"
                     /number=6
     exon            644..771
                     /gene="HUMTCHIT"
                     /number=7
     exon            772..954
                     /gene="HUMTCHIT"
                     /number=8
     exon            955..1071
                     /gene="HUMTCHIT"
                     /number=9
     exon            1072..1433
                     /gene="HUMTCHIT"
                     /number=10
     3'UTR           1210..1433
                     /gene="HUMTCHIT"
     polyA_signal    1407..1412
                     /gene="HUMTCHIT"
     polyA_site      1434
                     /gene="HUMTCHIT"
                     /note="32 A nucleotides"
BASE COUNT      375 a    360 c    342 g    357 t
```

```
ORIGIN

  Query Match             9.5%;  Score 157;  DB 74;  Length 1434;
  Best Local Similarity  62.9%;  Pred. No. 1.32e-125;
  Matches   710;  Conservative    0;  Mismatches 391;  Indels  27;  Gaps  10;

Db     69 ctgggcaggtgtagtggtcttgctgcttctccagggaggatctgcctacaaactggtttg 128
        |||||||||| |  ||||| |||||  |  |||    | || |||||    |||||||| ||
Qy     44 CTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTG 103

Db    129 ctactttaccaactggtcccaggaccggcaggaaccaggaaaattcacccctgagaatat 188
        |||||| |||||||||| |||||  || | ||||     |      |||   ||  || |  |
Qy    104 CTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTT 163

Db    189 tgacccctcctatgctctcatctcatctattcattcgccagcatcgaaaacaacaaggt 248
        ||||||  ||| ||| | || |||||||  | |||||  ||||    ||| || || |
Qy    164 GGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCT 223

Db    249 tatcatcaaggacaagagtgaagtgatgctctaccagaccatcaacagtctcaaaaccaa 308
        | || ||  ||   || ||| | ||  |||||||||    ||||  | || || |  |
Qy    224 GAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGAT 283

Db    309 gaatcccaaactgaaaattctcttgtccattggagggtacctgtttggttccaaagggtt 368
        |||||||||| ||||| |  || ||  |||| ||||| |      ||  |  |  |   |||
Qy    284 GAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCAGCACTCAGAAGTT 343

Db    369 ccaccctatggtggattcttctacatcacgcttggaattcattaactccataatcctgtt 428
        |    ||||||  |   |  | |     ||   |   ||  | |||||    |||  |||
Qy    344 CACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTT 403

Db    429 tctgaggaaccataactttgatggactggatgtaagctggatctacccag--atc-agaa 485
        |||| | ||  | | |||||| || || ||  |   ||||   |||||||  | | ||
Qy    404 TCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGCCAGGG 463

Db    486 -agaa-------a-acactcat---ttcactgtgctgattcatgagttagcagaagcctt 533
           ||         | |||    |    |||||    ||| | || || ||  ||  | |||||
Qy    464 GAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAATGCCTT 523

Db    534 tcagaaggacttcacaaaatccaccaaggaaaggcttctcttgactgcgggcgtatctgc 593
        ||| ||||   |   |  ||    |||||| | |||||  ||| ||| |  ||  | ||
Qy    524 CCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCCAGC 583

Db    594 agggaggcaaatgattgataacagctatcaagttgagaaactggcaaaagatctggattt 653
        ||| |       | |||    | ||  | || || ||| | ||  |  | ||||||||
Qy    584 TGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATTT 643

Db    654 catcaacctcctgtcctttgacttccatgggtcttgggaaaagccccttatcactggcca 713
        |||||||  || |||  |||||||||||| |||||||| |||       |||| || ||
Qy    644 TGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGG------TCACGGGACA 697

Db    714 caacagccctctgagcaaggggtggcaggacagagggccaagctcctactacaatgtgga 773
        |||||||| ||   |||| ||    | || || ||  ||    ||  |  ||| |||||
Qy    698 TAACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGA 757

Db    774 atatgctgtggggtactggatacataagggaatgccatcagagaaggtggtcatgggcat 833
           |||||||    | ||| | || |||||  |  ||  |    ||| || || | |||||
Qy    758 TGCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCAT 817

Db    834 ccccacatatgggcactccttcacactggcctc-tg--cagaaaccaccgtgggggcccc 890
           || || || || | |||||||||||||||||| |   |||| ||||  ||||||||||
Qy    818 GCCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCC 877

Db    891 tgcctctggccctggagctgctggacccatcacagagtcttcaggcttcctggcctatta 950
        ||| | ||  ||||  || | || ||| ||||  ||      |||  | ||||||| ||
Qy    878 AGCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTA 937

Db    951 tgagatctgccagttcctgaaaggagccaagatcacgcggctccaggatcagcaggttcc 1010
        |||  |||||    | |  ||| || ||||  |    | |  ||||||||||| |||| ||
Qy    938 TGAAGTCTGC---TCCTGGAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCC 994

Db   1011 ctacgcagtcaaggggaaccagtgggtgggctatgatgatgtgaagagtatggagaccaa 1070
        ||||    ||  ||  |||||||||||||||| ||||||||| ||||  |  | |||||
Qy    995 CTACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAA 1054
```

```
Db   1071 ggttcagttcttaaagaatttaaacctgggaggagccatgatctggtctattgacatgga 1130
          |||    |   | ||| |       ||||| || |||||| ||||| |  | ||| | ||
Qy   1055 GGTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGA 1114

Db   1131 tgacttcactggcaaatcctgcaaccagggcccttaccctcttgtcca 1178
          ||||||  | |||   ||||||||||||||||  ||||| ||  ||||
Qy   1115 TGACTTTGCCGGCTTCTCCTGCAACCAGGGCCGATACCCCCTCATCCA 1162


RESULT    5
LOCUS       HSU58515     1500 bp    mRNA            PRI       24-JUL-1996
DEFINITION  Human chitinase (HUMTCHIT) mRNA, exon 1b form, partial cds.
ACCESSION   U58515
NID         g1439567
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1500)
  AUTHORS   Grossman,A., Matsuyama,T., Baker,E., Waterhouse,P., Sutherland,G.R.
            and Mak,T.W.
  TITLE     Cloning of a novel lymphoid restricted human chitinase and
            localization to 1p13.3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1500)
  AUTHORS   Grossman,A., Matsuyama,T., Baker,E., Waterhouse,P., Sutherland,G.R.
            and Mak,T.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1996) Medical Biophysics, Ontario Cancer
            Institute, Princess Margaret Hospital, 610 University Avenue,
            Toronto, Ontario M5G 2M9, Canada
FEATURES             Location/Qualifiers
     source          1..1500
                     /organism="Homo sapiens"
                     /map="1p13.3"
                     /chromosome="1"
                     /clone_lib="Stratgene Jurkat T-cell cDNA library in Lambda
                     Zap II; Clontech thymus polyA+ RNA; Stratagene placental
                     genomic library in Lambda FIX II; activated T-cell plasmid
                     library (Dr. Brian Seed)"
     exon            1..142
                     /gene="HUMTCHIT"
                     /note="exon 1b; second alternatively spliced exon 1"
                     /number=1
     CDS             <1..1275
                     /evidence=not_experimental
                     /gene="HUMTCHIT"
                     /note="submitter can find no appropriate Kozak initiator
                     methionine, despite the open reading frame continuing
                     upstream, and is certain that there are no other exons 5'
                     as determined by RACE and genomic sequencing; use of a Met
                     further downstream as the initiator Met would result in
                     the deletion of amino acids conserved in all other family
                     members; submitter hypothesizes that an Ile is the
                     initiator codon for this form; second alternatively
                     spliced form; see GenBank Accession Number U58514 for the
                     first alternatively spliced form"
                     /codon_start=1
                     /product="chitinase"
                     /db_xref="PID:g1439568"
                     /translation="SQQLWLGSPDEVWLYLVCEHTHIFTAIIWDSRTRFGSTDFSFHP
                     SIAGVVVLLLLQGGSAYKLVCYFTNWSQDRQEPGKFTPENIDPFLCSHLIYSFASIEN
                     NKVIIKDKSEVMLYQTINSLKTKNPKLKILLSIGGYLFGSKGFHPMVDSSTSRLEFIN
                     SIILFLRNHNFDGLDVSWIYPDQKENTHFTVLIHELAEAFQKDFTKSTKERLLLTAGV
                     SAGRQMIDNSYQVEKLAKDLDFINLLSFDFHGSWEKPLITGHNSPLSKGWQDRGPSSY
                     YNVEYAVGYWIHKGMPSEKVVMGIPTYGHSFTLASAETTVGAPASGPGAAGPITESSG
                     FLAYYEICQFLKGAKITRLQDQQVPYAVKGNQWVGYDDVKSMETKVQFLKNLNLGGAM
                     IWSIDMDDFTGKSCNQGPYPLVQAVKRSLGSL"
     exon            143..172
                     /gene="HUMTCHIT"
                     /number=2
     exon            173..375
                     /gene="HUMTCHIT"
                     /number=3
     exon            376..431
                     /gene="HUMTCHIT"
                     /number=4
     exon            432..582
                     /gene="HUMTCHIT"
                     /number=5
     exon            583..709
                     /gene="HUMTCHIT"
                     /number=6
     exon            710..847
                     /gene="HUMTCHIT"
                     /number=7
     exon            848..1020
                     /gene="HUMTCHIT"
                     /number=8
     exon            1021..1137
                     /gene="HUMTCHIT"
                     /number=9
     exon            1138..1277
                     /gene="HUMTCHIT"
                     /number=10
     3'UTR           1276..1499
                     /gene="HUMTCHIT"
     exon            1278..1499
                     /gene="HUMTCHIT"
                     /number=11
     polyA_signal    1473..1478
                     /gene="HUMTCHIT"
     polyA_site      1500
                     /gene="HUMTCHIT"
                     /note="18 A nucleotides"
BASE COUNT      385 a    380 c    351 g    384 t
ORIGIN

  Query Match             9.4%;  Score 155;  DB 74;  Length 1500;
  Best Local Similarity  62.9%;  Pred. No. 1.56e-123;
  Matches   707;  Conservative    0;  Mismatches 390;  Indels  27;  Gaps  10;

Db    139 gcaggtgtagtggtcttgctgcttctccagggaggatctgcctacaaactggtttgctac 198
          |||||| |  ||||| ||||| |  |||    | || |||||    |||||||| ||||||
Qy     48 GCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTGCTAC 107

Db    199 tttaccaactggtcccaggaccggcaggaaccaggaaaattcacccctgagaatattgac 258
          || |||||||| ||||| || | ||||      |     |||   ||  || |  | |||
Qy    108 TTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGAC 167

Db    259 cccttcctatgctctcatctcatctattcattcgccagcatcgaaaacaacaaggttatc 318
          |||  ||| ||| | || |||||||  | |||||  ||||    ||| || || | | |
Qy    168 CCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCTGAGC 227

Db    319 atcaaggacaagagtgaagtgatgctctaccagaccatcaacagtctcaaaaccaagaat 378
          | ||  ||   || ||| | ||  ||||||||    |||| | || || |  | ||||
Qy    228 ACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGATGAAT 287

Db    379 cccaaactgaaaattctcttgtccattggagggtacctgtttggttccaaagggttccac 438
          ||||| ||||| |  || ||  |||| ||||| |     ||  |  |  |  ||||
Qy    288 CCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCAGCACTCAGAAGTTCACA 347

Db    439 cctatggtggattcttctacatcacgcttggaattcattaactccataatcctgtttctg 498
            |||||| |   |  | |     ||   |  ||  | |||||    |||  |||||||
Qy    348 GATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTTTCTG 407

Db    499 aggaaccataactttgatggactggatgtaagctggatctacccag--atc-agaa-aga 554
            | ||  | | |||||| || || ||  |   ||||   |||||||  | | ||   ||
Qy    408 CGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGCCAGGGGAGC 467

Db    555 a-------a-acactcat---ttcactgtgctgattcatgagttagcagaagcctttcag 603
                  | |||   |    |||||    ||| | || || || ||  | ||||| |||
Qy    468 CCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAATGCCTTCCAG 527
```

```
Db    604 aaggacttcacaaaatccaccaaggaaaggcttctcttgactgcgggcgtatctgcaggg 663
Qy    528 CAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCCAGCTGGG 587

Db    664 aggcaaatgattgataacagctatcaagttgagaaactggcaaaagatctggatttcatc 723
Qy    588 CAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATTTTGTC 647

Db    724 aacctcctgtcctttgacttccatgggtcttgggaaaagccccttatcactggccacaac 783
Qy    648 AACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGG------TCACGGGACATAAC 701

Db    784 agccctctgagcaaggggtggcaggacagagggccaagctcctactacaatgtggaatat 843
Qy    702 AGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGATGCT 761

Db    844 gctgtggggtactggatacataagggaatgccatcagagaaggtggtcatgggcatcccc 903
Qy    762 GCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATGCCT 821

Db    904 acatatggacactccttcacactggcctc-tg--cagaaaccaccgtgggggcccctgcc 960
Qy    822 ACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCAGCC 881

Db    961 tctggccctggagctgctggacccatcacagagtcttcaggcttcctggcctattatgag 1020
Qy    882 ACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTATGAA 941

Db   1021 atctgccagttcctgaaaggagccaagatcacgcggctccaggatcagcaggttccctac 1080
Qy    942 GTCTGC---TCCTGGAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCCTAC 998

Db   1081 gcagtcaaggggaaccagtgggtgggctatgatgatgtgaagagtatggagaccaaggtt 1140
Qy    999 ATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAGGTC 1058

Db   1141 cagttcttaaagaatttaaacctgggaggagccatgatctggtctattgacatggatgac 1200
Qy   1059 AGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGATGAC 1118

Db   1201 ttcactggcaaatcctgcaaccagggcccttaccctcttgtcca 1244
Qy   1119 TTTGCCGGCTTCTCCTGCAACCAGGGCCGATACCCCCTCATCCA 1162
```

```
RESULT    6
LOCUS       SSGP38KD     1733 bp    RNA             MAM       28-AUG-1996
DEFINITION  S.scrofa 38kDa heparin-binding glycoprotein.
ACCESSION   Z47803
NID         g634097
KEYWORDS    gp38k; heparin-binding glycoprotein.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Artiodactyla; Suiformes; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 1733)
  AUTHORS   Millis,A.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-1995) Millis A. J., University at Albany,
            Biological Sciences, 1400 Washington Avenue, Albany, NY, USA, 12222
REFERENCE   2  (bases 1 to 1733)
  AUTHORS   Shackelton,L.M., Mann,D.M. and Millis,A.J.
  TITLE     Identification of a 38-kDa heparin-binding glycoprotein (gp38k) in
            differentiating vascular smooth muscle cells as a member of a group
            of proteins associated with tissue remodeling
  JOURNAL   J. Biol. Chem. 270 (22), 13076-13083 (1995)
  MEDLINE   95286589
FEATURES             Location/Qualifiers
     source          1..1733
                     /organism="Sus scrofa"
                     /clone="pBS38k"
                     /dev_stage="adult"
                     /tissue_type="smooth muscle"
                     /clone_lib="lambda ZAP II"
     CDS             67..1218
                     /codon_start=1
                     /product="38 kDa heparin-binding glycoprotein"
                     /db_xref="PID:g634098"
                     /translation="MGLRVAQTGFVALVLLQSCAAYKLVCYYTSWSQYREGDGSCFPD
                     AIDPFLCTHIIYSFANISNNEIDTLEWNDVTLYDTLNTLKNRNPNLKTLLSVGGWNFG
                     SQRFSKIASNTQSRRTFIKSVPPFLRTHGFDGLDLAWISPGRRDKRHLTTLVKEMKAE
                     FVREALPGTERLLLSGAVSAGKVAIDRGYDIAQISQHLDFISLLTYDFHGAWRQTTGH
                     HSPLFRGQGDASSDRFSNADYAVSYVLRLGAPANKLVMGIPTFGRSFTLASSKTDVGA
                     PASGPGIPGRFTKEKGILAYYEICDFLQGATVRRPLGQQVPYATKGNQWVGYDDQESV
                     KNKAKYLKSRQLAGAMVWTLDLDDFRGNFCGQNLRFPLTSAIKDVLAAA"
BASE COUNT      408 a    507 c    463 g    355 t
ORIGIN
```

```
  Query Match             9.3%;  Score 154;  DB 48;  Length 1733;
  Best Local Similarity  62.5%;  Pred. No. 1.69e-122;
  Matches   687;  Conservative    0;  Mismatches 389;  Indels  24;  Gaps   9;

Db     79 gtggctcagacaggttttgtggccctggtgctgctccagagctgtgctgcctacaagctg 138
Qy     39 GTGGCCTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTG 98

Db    139 gtttgctactataccagctggtctcagtaccgggagggtgatgggagctgcttcccagat 198
Qy     99 GTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAG 158

Db    199 gccatcgacccccttcctctgcacccacatcatctacagctttgccaacataagcaacaat 258
Qy    159 GACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCAC 218

Db    259 gagattgacaccttggagtggaatgatgtgacgctctatgacacactgaacacactcaag 318
Qy    219 CAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAG 278

Db    319 aacaggaaccccaacctgaagaccctcctgtctgttggaggatggaactttggttctcaa 378
Qy    279 AAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCAGCACTCAG 338

Db    379 agattttccaaaatagcctccaacactcagagtcgcaggactttcatcaagtcggtgcca 438
Qy    339 AAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATC 398

Db    439 ccattcctgcggacccatggctttgatggactggacctagcctggatctcccctgg--gc 496
Qy    399 AGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGC 458

Db    497 ----ggag---------agacaagcggcatctcaccactctagtcaaggagatgaaggct 543
Qy    459 CAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAAT 518

Db    544 gagtttgtaagggaagccctg-cct--ggaacagagcggctcctgctcagcggagcggtg 600
Qy    519 GCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTT 578

Db    601 tctgcagggaaggtcgccattgacagaggctatgacatcgcccagatatctcaacacctg 660
Qy    579 CCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTG 638

Db    661 gacttcatcagccttttgacctatgacttccacggtgcctggcgccaaactacaggacac 720
Qy    639 GATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACAT 698

Db    721 cacagtcccctgttccgaggccaggggatgcaagttccgacagattcagcaatgcagac 780
Qy    699 AACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGAT 758

Db    781 tatgctgtgagctatgtgctgaggctgggggccccagccaataagctggtgatgggtatc 840
Qy    759 GCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATG 818

Db    841 cccacttttgggaggagcttcactttggcctcttcca-agacag-ac-gtgggagccccg 897
```

```
Qy    819 CCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCA 878

Db    898 gcatcagggccaggaataccaggccggttcaccaaggagaaagggatccttgcctactat 957
Qy    879 GCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTAT 938

Db    958 gagatctgtgacttcctccaaggagccacagtccgtagacccccttggccagcaggtcccc 1017
Qy    939 GAAGTCTG---CTCCTGGAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCC 995

Db   1018 tatgccaccaagggcaaccagtgggtggggtatgatgaccaggagagcgtcaaaaacaag 1077
Qy    996 TACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAG 1055

Db   1078 gcaaagtacctgaagagcaggcagctggctggtgctatggtgtggaccctggacttggat 1137
Qy   1056 GTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGAT 1115

Db   1138 gactttcggggcaacttctg 1157
Qy   1116 GACTTTGCCGGCTTCTCCTG 1135
```

```
RESULT    7
LOCUS       SSU19900     1733 bp    mRNA            MAM       02-FEB-1996
DEFINITION  Sus scrofa 38kDa heparin-binding glycoprotein mRNA, complete cds.
ACCESSION   U19900
NID         g643470
KEYWORDS    .
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Artiodactyla; Suiformes; Suina; Suidae; Sus.
REFERENCE   1  (sites)
  AUTHORS   Shackelton,L.M., Mann,D.M. and Millis,A.J.
  TITLE     Identification of a 38-kDa heparin-binding glycoprotein (gp38k) in
            differentiating vascular smooth muscle cells as a member of a group
            of proteins associated with tissue remodeling
  JOURNAL   J. Biol. Chem. 270 (22), 13076-13083 (1995)
  MEDLINE   95286589
REFERENCE   2  (bases 1 to 1733)
  AUTHORS   Millis,A.J.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JAN-1995) Albert J.T. Millis, Biological Sciences,
            University at Albany, 1400 Washington Avenue, Albany, NY 12222, USA
FEATURES             Location/Qualifiers
     source          1..1733
                     /organism="Sus scrofa"
                     /note="in Lambda Zap II"
                     /clone="pBS38k"
                     /tissue_type="smooth muscle"
                     /dev_stage="adult"
     CDS             67..1218
                     /note="gp38k"
                     /codon_start=1
                     /product="38kDa heparin-binding glycoprotein"
                     /db_xref="PID:g643471"
                     /translation="MGLRVAQTGFVALVLLQSCAAYKLVCYYTSWSQYREGDGSCFPD
                     AIDPFLCTHIIYSFANISNNEIDTLEWNDVTLYDTLNTLKNRNPNLKTLLSVGGWNFG
                     SQRFSKIASNTQSRRTFIKSVPPFLRTHGFDGLDLAWISPGRRDKRHLTTLVKEMKAE
                     FVREALPGTERLLLSGAVSAGKVAIDRGYDIAQISQHLDFISLLTYDFHGAWRQTTGH
                     HSPLFRGQGDASSDRFSNADYAVSYVLRLGAPANKLVMGIPTFGRSFTLASSKTDVGA
                     PASGPGIPGRFTKEKGILAYYEICDFLQGATVRRPLGQQVPYATKGNQWVGYDDQESV
                     KNKAKYLKSRQLAGAMVWTLDLDDFRGNFCGQNLRFPLTSAIKDVLAAA"
BASE COUNT      408 a    507 c    463 g    355 t
ORIGIN
```

```
  Query Match             9.3%;  Score 154;  DB 48;  Length 1733;
  Best Local Similarity  62.5%;  Pred. No. 1.69e-122;
  Matches   687;  Conservative    0;  Mismatches 389;  Indels  24;  Gaps   9;

Db     79 gtggctcagacaggttttgtggccctggtgctgctccagagctgtgctgcctacaagctg 138
```

```
Qy     39 GTGGCCTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTG 98

Db    139 gtttgctactataccagctggtctcagtaccgggagggtgatgggagctgcttcccagat 198
Qy     99 GTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAG 158

Db    199 gccatcgacccccttcctctgcacccacatcatctacagctttgccaacataagcaacaat 258
Qy    159 GACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCAC 218

Db    259 gagattgacaccttggagtggaatgatgtgacgctctatgacacactgaacacactcaag 318
Qy    219 CAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAG 278

Db    319 aacaggaaccccaacctgaagaccctcctgtctgttggaggatggaactttggttctcaa 378
Qy    279 AAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCAGCACTCAG 338

Db    379 agattttccaaaatagcctccaacactcagagtcgcaggactttcatcaagtcggtgcca 438
Qy    339 AAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATC 398

Db    439 ccattcctgcggacccatggctttgatggactggacctagcctggatctcccctgg--gc 496
Qy    399 AGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGC 458

Db    497 ----ggag---------agacaagcggcatctcaccactctagtcaaggagatgaaggct 543
Qy    459 CAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAAT 518

Db    544 gagtttgtaagggaagccctg-cct--ggaacagagcggctcctgctcagcggagcggtg 600
Qy    519 GCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTT 578

Db    601 tctgcagggaaggtcgccattgacagaggctatgacatcgcccagatatctcaacacctg 660
Qy    579 CCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTG 638

Db    661 gacttcatcagccttttgacctatgacttccacggtgcctggcgccaaactacaggacac 720
Qy    639 GATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACAT 698

Db    721 cacagtcccctgttccgaggccaggggatgcaagttccgacagattcagcaatgcagac 780
Qy    699 AACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGAT 758

Db    781 tatgctgtgagctatgtgctgaggctgggggccccagccaataagctggtgatgggtatc 840
Qy    759 GCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATG 818

Db    841 cccacttttgggaggagcttcactttggcctcttcca-agacag-ac-gtgggagccccg 897
Qy    819 CCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCA 878

Db    898 gcatcagggccaggaataccaggccggttcaccaaggagaaagggatccttgcctactat 957
Qy    879 GCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTAT 938

Db    958 gagatctgtgacttcctccaaggagccacagtccgtagacccccttggccagcaggtcccc 1017
Qy    939 GAAGTCTG---CTCCTGGAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCC 995

Db   1018 tatgccaccaagggcaaccagtgggtggggtatgatgaccaggagagcgtcaaaaacaag 1077
Qy    996 TACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAG 1055

Db   1078 gcaaagtacctgaagagcaggcagctggctggtgctatggtgtggaccctggacttggat 1137
Qy   1056 GTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGAT 1115

Db   1138 gactttcggggcaacttctg 1157
Qy   1116 GACTTTGCCGGCTTCTCCTG 1135
```

```
RESULT    8
ID   SSGP38KD   standard; RNA; MAM; 1733 BP.
AC   Z47803;
NI   g634097
DT   19-JAN-1995 (Rel. 42, Created)
DT   28-AUG-1996 (Rel. 49, Last updated, Version 6)
DE   S.scrofa 38kDa heparin-binding glycoprotein.
KW   gp38k; heparin-binding glycoprotein.
OS   Sus scrofa (domestic pig)
OC   Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC   Theria; Eutheria; Artiodactyla; Suiformes; Suidae.
RN   [2]
RP   1-1733
RA   Millis A.J.;
RT   ;
RL   Submitted (19-JAN-1995) to the EMBL/GenBank/DDBJ databases.
RL   Millis A. J., University at Albany, Biological Sciences, 1400
RL   Washington Avenue, Albany, NY, USA, 12222
RN   [3]
RP   1-1733
RX   MEDLINE; 95286589.
RA   Shackelton L.M., Mann D.M., Millis A.J.;
RT   "Identification of a 38-kDa heparin-binding glycoprotein (gp38k)
RT   in differentiating vascular smooth muscle cells as a member of a
RT   group of proteins associated with tissue remodeling";
RL   J. Biol. Chem. 270:13076-13083(1995).
FH   Key             Location/Qualifiers
FH
FT   source          1..1733
FT                   /organism="Sus scrofa"
FT                   /clone="pBS38k"
FT                   /dev_stage="adult"
FT                   /tissue_type="smooth muscle"
FT                   /clone_lib="lambda ZAP II"
FT   CDS             67..1218
FT                   /product="38 kDa heparin-binding glycoprotein"
FT                   /db_xref="PID:g634098"
FT                   /translation="MGLRVAQTGFVALVLLQSCAAYKLVCYYTSWSQYREGDGSCFPD
  A
FT                   IDPFLCTHIIYSFANISNNEIDTLEWNDVTLYDTLNTLKNRNPNLKTLLSVGGWNFGS
  Q
FT                   RFSKIASNTQSRRTFIKSVPPFLRTHGFDGLDLAWISPGRRDKRHLTTLVKEMKAEFV
  R
FT                   EALPGTERLLLSGAVSAGKVAIDRGYDIAQISQHLDFISLLTYDFHGAWRQTTGHHSP
  L
FT                   FRGQGDASSDRFSNADYAVSYVLRLGAPANKLVMGIPTFGRSFTLASSKTDVGAPASG
  P
FT                   GIPGRFTKEKGILAYYEICDFLQGATVRRPLGQQVPYATKGNQWVGYDDQESVKNKAK
  Y
FT                   LKSRQLAGAMVWTLDLDDFRGNFCGQNLRFPLTSAIKDVLAAA"
SQ   Sequence 1733 BP; 408 A; 507 C; 463 G; 355 T; 0 other;

  Query Match             9.3%;  Score 154;  DB 15;  Length 1733;
  Best Local Similarity  62.5%;  Pred. No. 1.69e-122;
  Matches   687;  Conservative    0;  Mismatches 389;  Indels  24;  Gaps   9;

Db      79 gtggctcagacaggttttgtggccctggtgctgctccagagctgtgctgcctacaagctg 138
           |||||   | |||||||  ||| |||| || || |||   |  |  ||||    || |||
Qy      39 GTGGCCTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTG 98

Db     139 gtttgctactataccagctggtctcagtaccgggagggtgatgggagctgcttcccagat 198
           || |||||||  |||| |||| | ||||||  |  |||| || |   ||| | | ||  |
Qy      99 GTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAG 158

Db     199 gccatcgacccccttcctctgcacccacatcatctacagctttgccaacataagcaacaat 258
           | | | ||||||  ||| |||||||||| |||||||  ||| ||   ||| | |||| |
Qy     159 GACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCAC 218

Db     259 gagattgacaccttggagtggaatgatgtgacgctctatgacacactgaacacactcaag 318
           || |   ||||  ||||||||||| | ||| |||||  |     | ||    || |||
Qy     219 CAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAG 278

Db     319 aacaggaaccccaacctgaagaccctcctgtctgttggaggatggaactttggttctcaa 378
           || | ||| ||||| ||||||||||  |  |  | |||||  |||||  ||  |  ||||
Qy     279 AAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCAGCACTCAG 338

Db     379 agattttccaaaatagcctccaacactcagagtcgcaggactttcatcaagtcggtgcca 438
           |  ||  |  | || |   |||   |  | |  ||   ||| ||  |||| ||||
Qy     339 AAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATC 398

Db     439 ccattcctgcggacccatggctttgatggactggacctagcctggatctcccctgg--gc 496
           || ||||| |   |  ||||||| || || ||||| | ||||   | ||| ||  ||
Qy     399 AGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGC 458

Db     497 ----ggag---------agacaagcggcatctcaccactctagtcaaggagatgaaggct 543
               ||||         |||||||  ||   |||| || || ||  ||||  ||     |
Qy     459 CAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAAT 518

Db     544 gagtttgtaagggaagccctg-cct--ggaacagagcggctcctgctcagcggagcggtg 600
           |  ||      |||||||| | |||  || |  || || || || || || || | ||||||
Qy     519 GCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTT 578

Db     601 tctgcagggaaggtcgccattgacagaggctatgacatcgcccagatatctcaacacctg 660
           | || ||| ||  |    | ||    || || ||  | | | | ||  | ||  |||||
Qy     579 CCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTG 638

Db     661 gacttcatcagccttttgacctatgacttccacggtgcctggcgccaaactacaggacac 720
           || ||  ||| |||| || |||| |||||||| ||  | |||    |    || |||||
Qy     639 GATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACAT 698

Db     721 cacagtcccctgttccgaggccaggggggatgcaagttccgacagattcagcaatgcagac 780
           |||| ||||| | |    | || |  ||      || | |       |  ||| |  ||
Qy     699 AACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGAT 758

Db     781 tatgctgtgagctatgtgctgaggctgggggccccagccaataagctggtgatgggtatc 840
           |||||||    |  |||| |  |||| |||| ||||  |||||| |  | || ||
Qy     759 GCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATG 818

Db     841 cccacttttgggaggagcttcactttggcctcttcca-agacag-ac-gtgggagccccg 897
           || || |  ||  |   ||||||  |||||||  ||   |||||  |  ||||| |||||
Qy     819 CCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCA 878

Db     898 gcatcagggccaggaataccaggccggttcaccaaggagaaagggatccttgcctactat 957
           ||  ||||| | || |  |||||||  ||||||||||   |||||| || |||||||||
Qy     879 GCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTAT 938

Db     958 gagatctgtgacttcctccaaggagccacagtccgtagacccccttggccagcaggtcccc 1017
           ||  ||||   || |   | || |||||    |  |||  ||  |  ||| |||| |||
Qy     939 GAAGTCTG---CTCCTGGAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCC 995

Db    1018 tatgccaccaagggcaaccagtgggtggggtatgatgaccaggagagcgtcaaaaacaag 1077
           ||   |  | || ||||||||||||||| | ||||||  |||||||  |||||| ||||
Qy     996 TACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAG 1055

Db    1078 gcaaagtacctgaagagcaggcagctggctggtgctatggtgtggaccctggacttggat 1137
           |  |  || ||||||   | |   ||||  || || ||||| ||| | |||||||| |||
Qy    1056 GTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGAT 1115

Db    1138 gactttcggggcaacttctg 1157
           ||||||   |||  || |||
Qy    1116 GACTTTGCCGGCTTCTCCTG 1135


RESULT    9
LOCUS       MMBRP39      1616 bp    RNA             ROD       28-NOV-1995
DEFINITION  M.musculus mRNA for BRP39 protein.
ACCESSION   X93035
NID         g1085065
KEYWORDS    brp39 gene; BRP39 protein.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 1616)
  AUTHORS   Morrison,B.W. and Leder,P.
  TITLE     neu and ras initiate murine mammary tumors that share genetic
            markers generally absent in c-myc and int-2-initiated tumors
  JOURNAL   Oncogene 9 (12), 3417-3426 (1994)
  MEDLINE   95060797
REFERENCE   2  (bases 1 to 1616)
  AUTHORS   Morrison,B.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-1995) B.W. Morrison, Dana-Farber Cancer
            Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1..1616
                     /organism="Mus musculus"
                     /strain="FUB/N"
                     /dev_stage="adult"
                     /tissue_type="breast"
                     /cell_line="SMF murine breast tumor"
                     /clone_lib="lambda ZAP II"
     misc_feature    66..128
                     /gene="brp39"
                     /note="leader sequence"
     CDS             66..1211
                     /gene="brp39"
                     /note="homology to glycosyl hydrolases"
                     /codon_start=1
                     /product="BRP39 protein"
                     /db_xref="PID:e211795"
                     /db_xref="PID:g1085066"
                     /translation="MGMRAALTGFAVLMLLQSCSAYKLVCYFTSWSQYREGVGSFLPD
                     AIQPFLCTHIIYSFANISSDNMLSTWEWNDESNYDKLNKLKTRNTNLKTLLSVGGWKF
                     GEKRFSEIASNTERRTAFVRSVAPFLRSYGFDGLDLAWLYPRLRDKQYFSTLIKELNA
                     EFTKEVQPGREKLLLSAALSAGKVAIDTGYDIAQIAQHLDFINLMTYDFHGVWRQITG
                     HHSPLFQGQKDTRFDRYSNVNYAVQYMIRLGAQASKLLMGIPTFGKSFTLASSENQLG
                     APISGEGLPGRFTKEAGTLAYYEICDFLKGAEVHRLSNEKVPFATKGNQWVGYEHKES
                     VKNKVGFLKEKKLAGAMVWALDLDDFQGTCQPKEFFPLTNAIKDALA"
BASE COUNT      420 a    442 c    404 g    350 t
ORIGIN

  Query Match             8.0%;  Score 133;  DB 81;  Length 1616;
  Best Local Similarity  62.4%;  Pred. No. 6.95e-101;
  Matches   673;  Conservative    0;  Mismatches 378;  Indels  27;  Gaps  13;

Db     88 caggctttgcggtcctgatgctgctccagagctgctctgcgtacaagctggtctgctact 147
Qy     49 CAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTGCTACT 108

Db    148 tcaccagctggtcccagtaccgggaaggcgttggaagcttcttaccagacgccatccaac 207
Qy    109 TCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACC 168

Db    208 ctttcctgtgcacccacatcatctacagctttgccaacatcagcagcgacaacatgctta 267
Qy    169 CCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCAC--CA-GCTGA 225

Db    268 gcacatgggagtggaatgacgagtcgaactatgacaagctgaataaactgaagaccagaa 327
Qy    226 GCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGATGA 285

Db    328 acaccaacctgaagaccctcctgtctgttggagggtggaaatttggcgaaaaaagatttt 387
Qy    286 ATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCAGCACTCAGAAGTTCA 345

Db    388 ccgagattgcctccaacactgagagacgcactgctttcgtccggtcggtagccccgttcc 447
Qy    346 CAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTTTC 405

Db    448 tgcgttcttatggctttgatgggctggatctcgcctggctctaccct------c------ 495
Qy    406 TGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGCCAGGGGA 465

Db    496 gct-taa-g-agacaagcagtatttctccaccctgatcaaggaactgaatgcggaattca 552
```

```
Qy    466 GCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAATGCCTTCC 525

Db    553 caaaggaggtccag--c-caggcagagagaaactcctgctcagcgcagctttgtcagcag 609
Qy    526 AGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCCAGCTG 585

Db    610 gaaaggtggccattgacactggctatgacatcgcccagatagcccaacacctggatttta 669
Qy    586 GGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATTTTG 645

Db    670 tcaatctcatgacctacgatttccatggagtctggcgccaaatcacaggccaccacagcc 729
Qy    646 TCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACATAACAGCC 705

Db    730 ccctcttccaaggccagaaggacactaggtttgacagatacagcaatgtgaactatgccg 789
Qy    706 CCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGATGCTGCTG 765

Db    790 tgcagtacatgatacgtctgggagcccaggccagcaagctactgatgggcatccccacct 849
Qy    766 TGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATGCCTACCT 825

Db    850 ttgggaagagcttcactctggcatcttc-tgaaa-atcagc-tgggagctccaatctcag 906
Qy    826 ACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCAGCCACAG 885

Db    907 gggaaggattaccaggccggttcaccaaggaggcagggaccctggcctactacgagatat 966
Qy    886 GGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTATGAAGTCT 945

Db    967 gcgacttcctcaaaggagctgaagtacatcgactctccaacgagaaggttcccttcgcta 1026
Qy    946 GCTCCTGG---AAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCCTACATCT 1002

Db   1027 ccaagggcaaccagtgggtggggtatgagcacaaggagagtgtcaaaaacaaggttgggt 1086
Qy   1003 TCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAGGTCAGCT 1062

Db   1087 tcctgaaggagaagaagctggcaggagccatggtgtgggcactggatttggatgattt 1144
Qy   1063 ATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGATGACTT 1120
```

```
RESULT   10
ID   MMD757     standard; RNA; ROD; 1506 BP.
AC   D87757;
NI   g1545818
DT   15-SEP-1996 (Rel. 49, Created)
DT   15-SEP-1996 (Rel. 49, Last updated, Version 1)
DE   House mouse; Musculus domesticus bone marrow cells mRNA for ECF-L
DE   precursor, complete cds.
KW   ECF-L precursor.
OS   Mus musculus (mouse)
OC   Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC   Theria; Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
RN   [1]
RP   1-1506
RA   Owhashi M.;
RT   ;
RL   Submitted (06-SEP-1996) to the EMBL/GenBank/DDBJ databases.
RL   Makoto Owhashi, The University of Tokushima, Faculty of Integrated
RL   Arts and Sciences; 1-1 Minami-Johsanjima, Tokushima, Tokushima 770,
RL   Japan (E-mail:ohashi@ias.tokushima-u.ac.jp, Tel:0886-56-7261,
RL   Fax:0886-56-7298)
RN   [2]
RP   1-1506
RA   Owhashi M., Arita H., Taoka Y., Hayai N.;
RT   "Molecular cloning and characterization of T lymphocyte-derived
RT   eosinophil chemotactic factor (ECF-L)";
RL   Unpublished.
FH   Key             Location/Qualifiers
FH
FT   source          1..1506
```

```
FT                   /organism="Mus musculus"
FT                   /cell_type="bone marrow cells"
FT                   /tissue_type="bone marrow"
FT   CDS             1..1197
FT                   /gene="mECF-L"
FT                   /codon_start=1
FT                   /product="ECF-L precursor"
FT                   /db_xref="PID:g1545819"
FT                   /translation="MAKLILVTGLAILLNVQLGSSYQLMCYYTSWAKDRPIEGSFKPG
  N
FT                   IDPCLCTHLIYAFAGMQNNEITYTHEQDLRDYEALNGLKDKNTELKTLLAIGGWKFGP
  A
FT                   SFSAMVSTPQNRQIFIQSVIRFLRQYNFDGLNLDWQYPGSRGSPPKDKHLFSVLVKEM
  R
FT                   KAFEEESVEKDIPRLLLTSTGAGIIDVIKSGYKIPELSQSLDYIQVMTYDLHDPKDGY
  T
FT                   GENSPLYKSPYDIGKSADLNVDSIISYWKDHGAASEKLIVGFPAYGHTFILSDPSKTG
  I
FT                   GAPTISTGPPGKYTDESGLLAYYEVCTFLNEGATEVWDAPQEVPYAYQGNEWVGYDNV
  R
FT                   SFKLKAQWLKDNNLGGAVVWRLDMDDFSGSFCHQRHFPLTSTLKGDLNIHSASCKGPY
  "
SQ   Sequence 1506 BP; 422 A; 337 C; 335 G; 412 T; 0 other;
```

```
  Query Match             7.3%;  Score 121;  DB 20;  Length 1506;
  Best Local Similarity  57.1%;  Pred. No. 1.14e-88;
  Matches   609;  Conservative    0;  Mismatches 452;  Indels   6;  Gaps   4;

Db      70 ctgatgtgctactataccagttgggctaaggacaggccaatagaagggagtttcaaacct 129
           ||| | |||||||  ||||  |||||  || |||| |     || |   | |||   ||
Qy      96 CTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCC 155

Db     130 ggtaatattgacccctgcctgtgtactcacctgatctatgcctttgctggaatgcagaat 189
               |  | |||||| |||| || || ||||| ||||| ||||| ||||| |||   ||
Qy     156 AAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAAC 215

Db     190 aatgagatcacttacacacatgagcaagacttgcgtgactatgaagcattgaatggtctg 249
           |  || | |    |||  |  | | |||  | |  |||  | |  || |||| |||
Qy     216 CACCAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTG 275

Db     250 aaagacaagaacactgagctaaaaactctcctggccattggaggatggaagtttggacct 309
           ||  | | |||  |  |||| || || ||  | ||||| ||||| ||||| ||  |  ||
Qy     276 AAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCAGCACT 335

Db     310 gcctcgttcagtgccatggtctctactcctcagaaccgtcagatattcattcagtcagtt 369
               |||||  |  |||||  | ||  |  | ||||||||||  ||  |  | || |
Qy     336 CAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCC 395

Db     370 atcagattccttcgtcaatataactttgatggcctcaacctggactggcagtaccctggg 429
           |||||  || || ||  |||| | |||||| |||||  |||| |||||| |||||||  ||
Qy     396 ATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGA 455

Db     430 tctcgaggaagccctcctaaggacaaacatctcttcagtgttctggtgaaggaaatgcgt 489
              |  || |||||| |    |||||  | | |||||     |||||  ||||  ||
Qy     456 AGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCC 515

Db     490 aaagcttttgaggaagaatctgtggagaaagacattccaaggctgctactcacttccaca 549
           || || ||  || | ||| |   |     ||  |    | | || || || | | |  |
Qy     516 AATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCG 575

Db     550 ggagcaggaatcattgacgtaatcaagtctgggtacaagatccctgaactgtctcagtct 609
           |  |||          |   |  |  |||| ||| || |     || |  | |||
Qy     576 GTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAAC 635

Db     610 cttgactatattcaggtcatgacatatgatctccatgatcctaaggatggctacactgga 669
           || || | |  |  | ||| | || ||  ||||||   ||  |||      ||| |||
Qy     636 CTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGA 695

Db     670 gaaaatagtcccctctataaatctccatatgacattggaaagagtgctgatctcaatgtg 729
           |  || || |||||||| ||    | | | || | |||       ||    ||||| |||
Qy     696 CATAACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTG 755

Db     730 gattcaatcatttcctactggaaggaccatggagcagcttctgagaagctcattgtggga 789
```

```
           ||| |     |     | |||  | |  | ||  |  || |    ||||| ||  | ||
Qy     756 GATGCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGC 815

Db     790 tttccagcatatgggcataccttttatcctgagtgacccttctaagactggaattggtgcc 849
           | ||  | || || |   ||||  |  |||     | | ||  | ||  || | || |||
Qy     816 ATGCCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCC 875

Db     850 cctacaattagtactggcccaccaggaaagtacacagatgaatcaggactcctggcttac 909
           ||  | |    |  ||||| | |||||    | |||  | |||  |||  | ||||| |||
Qy     876 CCAGCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTAC 935

Db     910 tatgaggtttgtacatttctgaatgaaggagccactgaggtctgggatgccccccaggaa 969
           |||||  ||  ||  |  | ||| | |  || |||||  |     |           ||| |
Qy     936 TATGAAGTCTG--C--TCCTGGAAGG-GG-GCCACCAAACAGAGAATCCAGGATCAGAAG 989

Db     970 gtaccctatgcctatcagggtaatgagtgggttggttatgacaatgtcaggagcttcaag 1029
           || |||||   ||  | ||  ||  |||||||  || | |||  ||||  |||||||||
Qy     990 GTGCCCTACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAA 1049

Db    1030 ttgaaggctcagtggctcaaggacaacaatttaggaggtgccgtggtctggcgcctggac 1089
              ||||     |  || ||| | ||     | || || ||| |||||||||   ||||||
Qy    1050 ACCAAGGTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGAC 1109

Db    1090 atggatgacttcagtggttctttctgtcaccagagacatttccctct 1136
           | ||||||||    || |  | |||  |||| | |  | ||| ||
Qy    1110 TTAGATGACTTTGCCGGCTTCTCCTGCAACCAGGGCCGATACCCCCT 1156
```

```
RESULT   11
LOCUS       D87757       1506 bp    mRNA            ROD       13-SEP-1996
DEFINITION  House mouse; Musculus domesticus bone marrow cells mRNA for ECF-L
            precursor, complete cds.
ACCESSION   D87757
NID         g1545818
KEYWORDS    ECF-L precursor.
SOURCE      Mus musculus bone marrow cells cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 1506)
  AUTHORS   Owhashi,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-1996) to the DDBJ/EMBL/GenBank databases. Makoto
            Owhashi, The University of Tokushima, Faculty of Integrated Arts
            and Sciences; 1-1 Minami-Johsanjima, Tokushima, Tokushima 770,
            Japan (E-mail:ohashi@ias.tokushima-u.ac.jp, Tel:0886-56-7261,
            Fax:0886-56-7298)
REFERENCE   2  (bases 1 to 1506)
  AUTHORS   Owhashi,M., Arita,H., Taoka,Y. and Hayai,N.
  TITLE     Molecular cloning and characterization of T lymphocyte-derived
            eosinophil chemotactic factor (ECF-L)
  JOURNAL   Unpublished (1996)
FEATURES             Location/Qualifiers
     source          1..1506
                     /organism="Mus musculus"
                     /cell_type="bone marrow cells"
                     /tissue_type="bone marrow"
     CDS             1..1197
                     /gene="mECF-L"
                     /codon_start=1
                     /product="ECF-L precursor"
                     /db_xref="PID:g1545819"
                     /translation="MAKLILVTGLAILLNVQLGSSYQLMCYYTSWAKDRPIEGSFKPG
                     NIDPCLCTHLIYAFAGMQNNEITYTHEQDLRDYEALNGLKDKNTELKTLLAIGGWKFG
                     PASFSAMVSTPQNRQIFIQSVIRFLRQYNFDGLNLDWQYPGSRGSPPKDKHLFSVLVK
                     EMRKAFEEESVEKDIPRLLLTSTGAGIIDVIKSGYKIPELSQSLDYIQVMTYDLHDPK
                     DGYTGENSPLYKSPYDIGKSADLNVDSIISYWKDHGAASEKLIVGFPAYGHTFILSDP
                     SKTGIGAPTISTGPPGKYTDESGLLAYYEVCTFLNEGATEVWDAPQEVPYAYQGNEWV
                     GYDNVRSFKLKAQWLKDNNLGGAVVWRLDMDDFSGSFCHQRHFPLTSTLKGDLNIHSA
                     SCKGPY"
BASE COUNT      422 a    337 c    335 g    412 t
ORIGIN
```

```
  Query Match            7.3%;  Score 121;  DB 81;  Length 1506;
  Best Local Similarity  57.1%;  Pred. No. 1.14e-88;
  Matches   609;  Conservative    0;  Mismatches 452;  Indels   6;  Gaps   4;

Db     70 ctgatgtgctactataccagttgggctaaggacaggccaatagaagggagtttcaaacct 129
Qy     96 CTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCC 155

Db    130 ggtaatattgacccctgcctgtgtactcacctgatctatgcctttgctggaatgcagaat 189
Qy    156 AAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAAC 215

Db    190 aatgagatcacttacacacatgagcaagacttgcgtgactatgaagcattgaatggtctg 249
Qy    216 CACCAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTG 275

Db    250 aaagacaagaacactgagctaaaaactctcctggccattggaggatggaagtttggacct 309
Qy    276 AAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCAGCACT 335

Db    310 gcctcgttcagtgccatggtctctactcctcagaaccgtcagatattcattcagtcagtt 369
Qy    336 CAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCC 395

Db    370 atcagattccttcgtcaatataactttgatggcctcaacctggactggcagtaccctggg 429
Qy    396 ATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGA 455

Db    430 tctcgaggaagccctcctaaggacaaacatctcttcagtgttctggtgaaggaaatgcgt 489
Qy    456 AGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCC 515

Db    490 aaagcttttgaggaagaatctgtggagaaagacattccaaggctgctactcacttccaca 549
Qy    516 AATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCG 575

Db    550 ggagcaggaatcattgacgtaatcaagtctgggtacaagatccctgaactgtctcagtct 609
Qy    576 GTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAAC 635

Db    610 cttgactatattcaggtcatgacatatgatctccatgatcctaaggatggctacactgga 669
Qy    636 CTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGA 695

Db    670 gaaaatagtcccctctataaatctccatatgacattggaaagagtgctgatctcaatgtg 729
Qy    696 CATAACAGCCCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTG 755

Db    730 gattcaatcatttcctactggaaggaccatggagcagcttctgagaagctcattgtggga 789
Qy    756 GATGCTGCTGTGCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGC 815

Db    790 tttccagcatatgggcataccttatcctgagtgacccttctaagactggaattggtgcc 849
Qy    816 ATGCCTACCTACGGACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCC 875

Db    850 cctacaattagtactggcccaccaggaaagtacacagatgaatcaggactcctggcttac 909
Qy    876 CCAGCCACAGGGTCTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTAC 935

Db    910 tatgaggtttgtacatttctgaatgaaggagccactgaggtctgggatgccccccaggaa 969
Qy    936 TATGAAGTCTG--C--TCCTGGAAGG-GG-GCCACCAAACAGAGAATCCAGGATCAGAAG 989

Db    970 gtaccctatgcctatcagggtaatgagtgggttggttatgacaatgtcaggagcttcaag 1029
Qy    990 GTGCCCTACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAA 1049

Db   1030 ttgaaggctcagtggctcaaggacaacaatttaggaggtgccgtggtctggcgcctggac 1089
Qy   1050 ACCAAGGTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGAC 1109

Db   1090 atggatgacttcagtggttctttctgtcaccagagacatttccctct 1136
Qy   1110 TTAGATGACTTTGCCGGCTTCTCCTGCAACCAGGGCCGATACCCCCT 1156
```

```
RESULT   12
LOCUS       SSU43490     2022 bp    mRNA            MAM       20-JAN-1996
DEFINITION  Sus scrofa oviductal secretory glycoprotein precursor mRNA,
            complete cds.
ACCESSION   U43490
NID         g1155331
KEYWORDS    .
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Artiodactyla; Suiformes; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 2022)
  AUTHORS   Buhi,W.C., Choi,I., Cleaver,B.D. and Simmen,F.A.
  TITLE     Molecular cloning and characterization of an estrogen-dependent
            porcine oviductal secretory glycoprotein (POSP)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2022)
  AUTHORS   Buhi,W.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-1995) William C. Buhi, University of Florida,
            Obstetrics and Gynecology, 1600 S.W. Archer Road, Gainesville, FL
            32610-0294, USA
FEATURES             Location/Qualifiers
     source          1..2022
                     /organism="Sus scrofa"
                     /strain="Crossbred Yorkshire x Duroc x Hampshire"
                     /clone="alv4"
                     /clone_lib="pig oviduct cDNA library of William C.Buhi"
                     /sex="female"
                     /cell_type="epithelial"
                     /tissue_type="oviduct"
                     /dev_stage="adult"
     5'UTR           <1..10
     CDS             11..1594
                     /codon_start=1
                     /product="oviductal secretory glycoprotein precursor"
                     /db_xref="PID:g1155332"
                     /translation="MGKLLLWVGLVLVLKHHNGAAHKLVCYFANWAFSRPGPASILPR
                     DLDPFLCTHLVFAFASMNDSQIVAKDARDESIFYPEFNQLKERNEKLKTLLSIGGWNF
                     GTSRFTTMLSTFTNREKFIRSAIGLLRTHGFDGLDLFFLYPGLRGSPRRDRWNFLFLL
                     EELLLAFRREAQLTMRPRLLLSAAVSADPHVIQKAYDVRLLGRLLDFINVLSYDLHGS
                     WEKVTGHNSPLFSLSDDPKSSAYTMNYWRKLGAPPEKLLMGFPTYGRTFRLLKASKNE
                     LGAEAVGPASPGKYTKQAGFLAYYEVCSFVQRAKKRWIDHQYVPYAYRGKEWVGYDDD
                     ISFSYKAFFIKKEHFGGAMVWTLDLDDVRGTFCGTGPFPLVYMLNDLLLKAEVSSTLS
                     PGFGLSTTVNSSRTCPESLAVTKDLTTDLGILPLGGEAVATETHGRSDNMTVTPGGGL
                     VAPTRPTLSFGKLTVAPEGKTESPGEKAMTPVGHPSVTPGDMSVPPVPIQTGDRITPP
                     RRQAVAPEKMTLPSGKRSD"
     sig_peptide     11..73
     mat_peptide     74..1591
                     /product="oviductal secretory glycoprotein"
     3'UTR           1595..2022
     polyA_signal    1987..1992
BASE COUNT      449 a    552 c    534 g    487 t
ORIGIN
```

```
  Query Match            6.2%;  Score 103;  DB 48;  Length 2022;
  Best Local Similarity  59.5%;  Pred. No. 1.38e-70;
  Matches   369;  Conservative    0;  Mismatches 248;  Indels   3;  Gaps   3;

Db     77 aagctggtgtgttattttgccaactgggcattcagtcggccaggccctgcttcgatcctg 136
Qy     93 AAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTG 152

Db    137 ccccgggacctggacccctttctctgcacccacctggtcttcgccttcgcctcaatgaac 196
Qy    153 CCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACC 212

Db    197 gacagtcagattgtggccaaggatgcccgggatgagagcatcttctacccagagttcaac 256
```

```
Qy    213 AACCACCAGCTGAGCACCACTGAGTGGAATGACGAGA-C-TCT-CTACCAGGAGTTCAAT 269

Db    257 cagctcaaggagaggaatgagaagctgaaaacactgctatccatcggagggtggaacttt 316

Qy    270 GGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTC 329

Db    317 ggcacatccaggttcaccacgatgctgtccaccttcaccaaccgggaaaagttcatcagg 376

Qy    330 AGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAAC 389

Db    377 tctgccataggcctcctgaggacacatggcttcgacgggctggacctcttcttcttgtac 436

Qy    390 TCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTAC 449

Db    437 cctggactcagaggcagccccaggcgtgaccgctggaactttctcttcttacttgaagag 496

Qy    450 CCAGGAAGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGAC 509

Db    497 ctcctgcttgccttcaggagggaggcacagctcaccatgcgtccgaggctgctgctctct 556

Qy    510 TTGGCCAATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGT 569

Db    557 gctgctgtctctgcggacccgcatgtcatccaaaaagcctatgatgtgcgccttctggga 616

Qy    570 GCAGCGGTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCC 629

Db    617 agactcctggatttcatcaatgtcttgtcttatgacttacatggaagctgggagaaggtc 676

Qy    630 CAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTC 689

Db    677 acgggacacaacagcccccct 696

Qy    690 ACGGGACATAACAGCCCCCT 709
```

```
RESULT   13
LOCUS       OAU16719     2034 bp    mRNA            MAM       21-JUL-1995
DEFINITION  Ovis aries estrogen dependent oviduct protein precursor mRNA,
            complete cds.
ACCESSION   U16719
NID         g885600
KEYWORDS    .
SOURCE      sheep.
  ORGANISM  Ovis aries
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Caprinae; Ovis.
REFERENCE   1  (bases 1 to 2034)
  AUTHORS   DeSouza,M.M. and Murray,M.K.
  TITLE     An estrogen-dependent secretory protein, which shares identity with
            chitinases, is expressed in a temporally and regionally specific
            manner in the sheep oviduct at the time of fertilization and embryo
            development
  JOURNAL   Endocrinology 136 (6), 2485-2496 (1995)
  MEDLINE   95269691
REFERENCE   2  (bases 1 to 2034)
  AUTHORS   DeSouza,M.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1994) Mary M. DeSouza, Tufts University, Anatomy
            and Cell Biology, 136 Harrison Ave, Boston, MA 02111, USA
FEATURES             Location/Qualifiers
     source          1..2034
                     /organism="Ovis aries"
                     /clone="MD2A, MD34, MD42"
                     /clone_lib="sheep oviduct cDNA library of M. Murray"
                     /sex="female"
                     /tissue_type="oviduct"
     5'UTR           1..16
     sig_peptide     17..78
     CDS             17..1636
                     /codon_start=1
                     /product="estrogen dependent oviduct protein precursor"
                     /db_xref="PID:g885601"
                     /translation="MGKLLLWVGLLLMLKHHDGAAHKLVCYFTNWAFSRPGSASILPR
                     DLDPFLCTHLVFAFASMNNNQIVPKDPLDEKILYPEFNKLKERNRGLKTLLSVGGWNF
                     GTSRFTKMLSTFSNRERFVKSVIALLRTHGFDGLDLFFLYPGLRGSPARDRWTFVFLL
                     EELLQAFKNEAQLTMRPRLLLSAAVSGDPHVIQKAYDARLLGRLLDFISVLSYDLHGS
                     WEKVTGHNSPLFSLPGDPKSSAYAMSYWRQLGVPPEKLLMGLPTYGRTFHLLRASQNE
                     LGAGAAGPASPGKYTKQAGFLAYYEVCSFVQRAKKRWINDQYVPYAFKGKEWVGYDDA
                     ISFGYKAFFIKREHFGGAMVWTLDLDDFRGNFCGTGPFPLAHTLNNLLVNDEFSSTPS
                     PKFWFSTAVNSSRIGPEMPTMTRDLTTGLGILPLGGEAVATETHRKSATMTTTPRGET
                     ATPTRTPLSSGRRTAAPEGKTESPGEKPLTSVGRLAVSPGGIAVGPVHLQIGQKVTPP
                     GRKAGVPEKVTIPSGKMTVTPDGRAETLERL"
     mat_peptide     79..1629
                     /note="secretory glycoprotein, Mr 90-92,000, fimbria and
                     ampulla specific"
                     /product="estrogen dependent oviduct protein"
     3'UTR           1637..2034
     polyA_signal    1855..1860
     polyA_signal    2012..2017
BASE COUNT      451 a    548 c    537 g    498 t
ORIGIN
```

```
  Query Match            5.7%;  Score 95;  DB 47;  Length 2034;
  Best Local Similarity  58.9%;  Pred. No. 1.15e-62;
  Matches   365;  Conservative    0;  Mismatches 252;  Indels   3;  Gaps   2;

Db     83 aagctggtgtgttatttcaccaactgggcattcagtcggcccggctctgcctcgatcctg 142

Qy     93 AAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTG 152

Db    143 ccccgggatctggaccccttttctctgcacccacctggtatttgcctttgcctcgatgaac 202

Qy    153 CCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACC 212

Db    203 aacaatcagattgttcctaaggatcccctggatgagaaaatcctctacccagagttcaac 262

Qy    213 AACCACCAGCTGAGCACCACTGAGTGGAATGACGAGAC--TC-TCTACCAGGAGTTCAAT 269

Db    263 aagctcaaggagaggaacaggggctgaaaacactgctgtccgtcgggggtggaacttc 322

Qy    270 GGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTC 329

Db    323 ggcacctcgaggttcaccaagatgctgtccacgttttccaaccgggaaaggtttgtcaaa 382

Qy    330 AGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAAC 389

Db    383 tcagtgatcgccctcctgaggacacatggcttcgatggtctggacctcttcttcttgtac 442

Qy    390 TCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTAC 449

Db    443 cctggactcagaggcagccccgcgcgtgaccgctggacctttgtcttcttacttgaggag 502

Qy    450 CCAGGAAGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGAC 509

Db    503 ctcctgcaggccttcaagaatgaggcccagctcaccatgcgcccaaggctgctgctgtct 562

Qy    510 TTGGCCAATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGT 569

Db    563 gctgccgtctctggggacccccacgtcatccagaaagcatatgatgcacgccttctgggc 622

Qy    570 GCAGCGGTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCC 629

Db    623 agactcctggatttcatcagcgtcttgtcttacgacttacatggaagctgggaaaaggtc 682

Qy    630 CAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTC 689

Db    683 acaggacacaatagccctct 702

Qy    690 ACGGGACATAACAGCCCCCT 709
```

```
RESULT   14
LOCUS       OAU17988     1596 bp    mRNA            MAM       30-MAY-1996
DEFINITION  Ovis aries 92kD oestrus-associated oviduct glycoprotein mRNA,
```

```
            partial cds.
ACCESSION   U17988
NID         g602293
KEYWORDS    .
SOURCE      sheep.
  ORGANISM  Ovis aries
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Caprinae; Ovis.
REFERENCE   1  (bases 1 to 1596)
  AUTHORS   Marshall,J.T., Brownlee,A.G. and Nancarrow,C.D.
  TITLE     Cloning and sequencing of a cDNA encoding ovine oestrous
            glycoprotein
  JOURNAL   Reprod. Fert. Dev. 8 (2), 305-310 (1996)
REFERENCE   2  (bases 1 to 1596)
  AUTHORS   Marshall,J.T.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-1994) James T. A. Marshall, Division of Animal
            Production, Commonwealth Scientific and Industrial Research
            Organization, Locked Bag 1, Delivery Centre, Blacktown, New South
            Wales, 2148, Australia
FEATURES             Location/Qualifiers
     source          1..1596
                     /organism="Ovis aries"
                     /strain="merino"
                     /sex="female"
                     /cell_type="epithelial cell"
                     /tissue_type="ampulla oviduct"
                     /dev_stage="oestrus"
     CDS             <1..1596
                     /codon_start=1
                     /product="92kD oestrus-associated oviduct glycoprotein"
                     /db_xref="PID:g602294"
                     /translation="LLLVLKHHDGAAHKLVCYFTNWAFSRPGSASILPRDLDPFLCTH
                     LVFAFASMNNNQIVPKDPLDEKILYPEFNKLKERNRGLKTLLSVGGWNFGTSRFTKML
                     STFSNRERFVNSVIALLRTHGFDGLDLFFLYPGLRGSPARDRWTFVFLLEELLQAFKN
                     EAQLTMRPRLLLSAAVSGDPHVIQKAYDARLLGRLLDFISVLSYDLHGSWEKVTGHNS
                     PLFSLPGDPKSSAYAMSYWRQLGVPPEKLLMGLPTYGRTFHLLRASQNELGAGAVGPA
                     SPGKYTKQAGFLAYYEVCSFVQRAKKRWINDQYVPYAFKGKEWVGYDDAISFGYKAFF
                     IKREHFGGAMVWTLDLDDFRGNFCGTGPFPLVHTLNNLLVNDEFSSTPSPKFWFSTAV
                     NSSRIGPEMPTMTRDLTTGLGILPLGGEAVATETHRKSATMTTTPRGETATPTRTPLS
                     SGRRTAAPEGKTESPGEKPLTSVGHLAVSPGGIAVGPVHLQIGQKVTPPGRKAGVPEK
                     VTTPSGKMTVTPDGRAETLERRL"
     mRNA            <1..>1596
                     /product="92kD oestrus-associated oviduct glycoprotein"
     sig_peptide     <1..36
                     /note="inferred from bovine sequence: GenBank Accession
                     Number D16639"
                     /codon_start=1
     primer_bind     1..19
                     /note="derived from PCR primer based on bovine cDNA
                     sequence: GenBank Accession Number D16639"
     primer_bind     complement(1575..1596)
                     /note="derived from PCR primer based on bovine cDNA
                     sequence: GenBank Accession Number D16639"
BASE COUNT      352 a    440 c    430 g    374 t
ORIGIN

  Query Match             5.5%;  Score 91;  DB 47;  Length 1596;
  Best Local Similarity  58.5%;  Pred. No. 9.79e-59;
  Matches   363;  Conservative    0;  Mismatches 254;  Indels   3;  Gaps   2;

Db     40 aagctggtgtgttatttcaccaactgggcattcagtcggcccggctctgcctcgatcctg 99
          || |||||  || || ||||||||||||||        | |  ||    ||     |||||
Qy     93 AAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTG 152

Db    100 ccccgggatctggaccccttttctctgcacccacctggtatttgcctttgcctcgatgaac 159
          |||  |||  ||||||||   || ||||||||||  | |  ||||| ||    |||| |
Qy    153 CCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACC 212

Db    160 aacaatcagattgttcctaaggatcccctggatgagaaaatcctctacccagagttcaac 219
          ||| | ||| |    | |  ||       || ||||   || ||||||  ||||||||
Qy    213 AACCACCAGCTGAGCACCACTGAGTGGAATGACGAGAC--TC-TCTACCAGGAGTTCAAT 269

Db    220 aagctcaaggagaggaacaggggctgaaaacactgctgtccgttgggggtggaacttc 279
          || ||| ||| |||      |||||| || ||| |  || | || || ||||| |||
Qy    270 GGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTC 329

Db    280 ggcacctcgaggttcaccaagatgctgtccacgttttccaaccgggaaaggtttgtcaat 339
          ||||   || ||||||  | ||| |  |||||     ||||||  | |  ||||||||
Qy    330 AGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAAC 389

Db    340 tcagtgatcgccctcctgaggacacatggcttcgatggtctggacctcttcttcttgtac 399
          || |  |||    | ||| | | | |  |||| || || || |||||   ||    ||||
Qy    390 TCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTAC 449

Db    400 cctggactcagaggcagccccgcgcgtgaccgctggacctttgtcttcttacttgaggag 459
          || |||  |   || ||||| ||    |||     |  |||       | |  |  ||||
Qy    450 CCAGGAAGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGAC 509

Db    460 ctcctgcaggccttcaagaatgaggcccagctcaccatgcgcccaaggctgctgctgtct 519
          |     | |||||| || | || ||||||  | |   |     | | || || |||  |
Qy    510 TTGGCCAATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGT 569

Db    520 gctgccgtctctggggaccccacgtcatccagaaagcatatgatgcacgccttctgggc 579
          || || ||  | |  |  |    |    | |    | ||| || |    |    | | |
Qy    570 GCAGCGGTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCC 629

Db    580 agactcctggatttcatcagcgtcttgtcttacgacttacacggaagctgggaaaaggtc 639
          |||||||||  ||| | |  || | |||||||| || ||    ||||| ||||||
Qy    630 CAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTC 689

Db    640 acaggacacaatagccctct 659
          || ||||| || |||||  ||
Qy    690 ACGGGACATAACAGCCCCCT 709


RESULT   15
LOCUS       BABEDOSG     2228 bp    mRNA            PRI       11-OCT-1996
DEFINITION  Papio hamadryas anubis estradiol-dependent oviduct-specific
            glycoprotein mRNA, partial cds.
ACCESSION   M59903
NID         g1256950
KEYWORDS    estradiol-dependent oviduct-specific glycoprotein.
SOURCE      Papio hamadryas anubis cDNA to mRNA.
  ORGANISM  Papio hamadryas anubis
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 2228)
  AUTHORS   Donnelly,K.M., Fazleabas,A.T., Verhage,H.G., Mavrogianis,P.A. and
            Jaffe,R.C.
  TITLE     Cloning of a recombinant complementary DNA to a baboon (Papio
            anubis) estradiol-dependent oviduct-specific glycoprotein
  JOURNAL   Mol. Endocrinol. 5 (3), 356-364 (1991)
  MEDLINE   91367180
REFERENCE   2  (bases 1 to 2228)
  AUTHORS   Jaffe,R.C., Arias,E.B., O'Day-Bowman,M.B., Donnelly,K.M.,
            Mavrogianis,P.A. and Verhage,H.G.
  TITLE     Regional Distribution and Hormonal Control of Estrogen-Dependent
            Oviduct Specific Glycoprotein mRNA in the Baboon (Papio anubis)
  JOURNAL   Biol. Reprod. (1996) In press
REFERENCE   3  (bases 1 to 2228)
  AUTHORS   Jaffe,R.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1991) Randal C. Jaffe, Department of Physiology
            and Biophysics, University of Illinois College of Medicine,
            Chicago, IL 60612-7342, USA
FEATURES             Location/Qualifiers
     source          1..2228
                     /organism="Papio hamadryas anubis"
     CDS             13..1884
                     /codon_start=1
                     /product="estradiol-dependent oviduct-specific
                     glycoprotein"
                     /db_xref="PID:g1256951"
```

```
                     /translation="MWKLLLWVGLVLVLKHHDGAAHKLVCYFTNWAHSRPGPASILPH
                     DLDPFLCTHLIFAFASMNNNQIVAKDLQDEKILYPEFNKLKERNRELKTLLSIGGWNF
                     GTSRFTTMLSTFANREKFIASVISLLRTHDFDGLDLFFLYPGLRGSPMHDRWTFLFLI
                     EELLFAFRKEALLTMRPRLLLSAAVSGVPHIVQTSYDVRFLGRLLDFINVLSYDLHGS
                     WEKFTGHNSPLFSLPEDPKSSAYAMNYWRKLGAPSEKLIMGIPTYGRTFRLLKASKNG
                     LQATAIGPASPGKYTKQAGFLAYFEICSFVWGAKKHWIDYQYVPYANKGKEWVGYDDA
                     ISFSYKAWFIRREHFGGAMVWTLDMDDVRGTFCGTGPFPLVYVMNDILVRAEFSSTSL
                     PQFWLSSAVNSSSTDPERLAVTKAWTTDIKILPPGGEAGVTEIHGKCENMTITPRVTI
                     VTPTKETVSLGKHTVALGEKTEITGATTMTSVGHQSMTPGEKALTPVGHQSELPGKKT
                     LTPVGHQSVTTGQKTLISVGYHSVTPGEKTLTPVGHPSVTPVSHQSVSPGGMTMTPVH
                     FQTETLRQNTMAPRRKAVAHEKVTVPSRKISVTPEGQTVPLRGEYLTSETGTHPQDG"
BASE COUNT      559 a    572 c    531 g    566 t
ORIGIN

  Query Match           5.5%;  Score 91;  DB 67;  Length 2228;
  Best Local Similarity  58.5%;  Pred. No. 9.79e-59;
  Matches   363;  Conservative    0;  Mismatches 254;  Indels   3;  Gaps   1;

Db      79 aaactggtgtgttatttcaccaactgggcacacagtcggccaggccccgcctcgatcttg 138
           |||||||| || || |||||||||||||| ||     | |  ||    ||     || ||
Qy      93 AAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTG 152

Db     139 ccccatgacctggaccccttttctctgcacccacctgatatttgcctttgcctcaatgaac 198
           ||| | ||| |||||||    || ||||||||||| || |  ||||| ||    |||| |
Qy     153 CCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACC 212

Db     199 aacaatcagattgttgctaaggatctccaggacgagaaaattctctacccagagttcaac 258
           ||| | ||| |     | |  ||     | |||||||    ||||||||  ||||||||
Qy     213 AACCACCAGCTGAGCACCACTGAGTGGAATGACGAGAC---TCTCTACCAGGAGTTCAAT 269

Db     259 aaactaaaggagaggaacagagagctgaaaacactgctgtccatcggcgggtggaacttt 318
             || ||| ||| |||     ||||||| || ||| |  |||||||  || ||||| ||
Qy     270 GGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTC 329

Db     319 ggcacctccaggttcaccactatgttgtccacatttgccaaccgtgaaaagtttattgct 378
            ||||    | ||||||   |||| |  ||||      |||||||  | |  ||| |
Qy     330 AGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAAC 389

Db     379 tcagttatatcccttctgaggacacatgactttgatggtcttgatctttttttcttatat 438
           || |  ||     ||||| | | | |   |||||| || ||||| |||    |     ||
Qy     390 TCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTAC 449

Db     439 cctggactaagaggcagccccatgcatgaccggtggacttttctcttcttaattgaagag 498
           || |||      || |||||       |||  |  |   ||       | |  |  | ||
Qy     450 CCAGGAAGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGAC 509

Db     499 ctcctgtttgccttccggaaggaggcactgctcaccatgcgcccaaggctgctgctgtct 558
            |      |||||||| | |||| || | |  | |   |     | | || || |||  |
Qy     510 TTGGCCAATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGT 569

Db     559 gctgctgtttctggggtcccacacatcgtccaaacatcctatgatgtgcgctttctagga 618
           || || ||| | |  |  |    |   ||  |  |    || || |||  |    | |
Qy     570 GCAGCGGTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCC 629

Db     619 agactcctggatttcatcaatgtcttgtcttatgacttacatggaagttgggaaaagttc 678
             |||||||||  ||||  |  || | || ||||| |||||   |||||| ||| ||
Qy     630 CAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTC 689

Db     679 acaggacataatagcccccct 698
           || |||||||| ||||||||
Qy     690 ACGGGACATAACAGCCCCCT 709


Search completed: Sun Jun 29 20:45:12 1997
Job time : 1558 secs.
```

****

FASTA (TM)

****

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Jun 29 20:45:32 1997; MasPar time 177.88 Seconds
861.257 Million cell updates/sec

Tabular output not generated.

Title: >US-08-663-618A-3
Description: (1-1656) from US08663618A.seq
Perfect Score: 1656
N.A. Sequence: 1 GCTGCAGCCTGCCGCTGAGC.....TGTGCCCCTCAAAAAAAAAA 1656
Comp: CGACGTCGGACGGCGACTCG.....ACACGGGGAGTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 121476 seqs, 46255616 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq26
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23

Statistics: Mean 9.577; Variance 5.573; scale 1.719

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	176	10.6	1681	13	Q85245	YKL-40 gene.	5.39e-101
	2	131	7.9	966	1	N81756	Gene encoding polypep	2.21e-69
	3	92	5.6	1047	2	Q10572	Human Natriuretic Pep	1.24e-42
	4	91	5.5	2504	15	Q90443	Murine oviduct specif	5.83e-42
	5	83	5.0	1994	15	Q90442	Bovine oviduct specif	1.33e-36
c	6	78	4.7	1047	2	Q10572	Human Natriuretic Pep	2.76e-33
	7	67	4.0	2366	15	Q90444	Hamster oviduct speci	4.37e-26
	8	51	3.1	204	1	N81164	Base substituted E.co	5.90e-16
c	9	45	2.7	91	9	Q51746	Oligonucleotide probe	2.64e-12
	10	42	2.5	91	9	Q51746	Oligonucleotide probe	1.60e-10
	11	38	2.3	114	12	Q70468	Generic DNA sequence	3.39e-08
c	12	38	2.3	204	1	N81164	Base substituted E.co	3.39e-08
	13	37	2.2	114	12	Q70467	Generic DNA sequence	1.26e-07
	14	37	2.2	114	12	Q70469	Generic DNA sequence	1.26e-07
	15	36	2.2	114	12	Q70465	Generic DNA sequence	4.64e-07
	16	36	2.2	114	12	Q70466	Generic DNA sequence	4.64e-07
	17	35	2.1	114	12	Q70473	Generic DNA sequence	1.69e-06
	18	34	2.1	114	12	Q70470	Generic DNA sequence	6.08e-06
	19	34	2.1	114	12	Q70472	Generic DNA sequence	6.08e-06
	20	34	2.1	114	12	Q70471	Generic DNA sequence	6.08e-06
c	21	34	2.1	114	12	Q70469	Generic DNA sequence	6.08e-06
c	22	33	2.0	114	12	Q70466	Generic DNA sequence	2.16e-05
c	23	33	2.0	114	12	Q70467	Generic DNA sequence	2.16e-05
c	24	32	1.9	114	12	Q70465	Generic DNA sequence	7.57e-05
c	25	32	1.9	114	12	Q70468	Generic DNA sequence	7.57e-05
c	26	31	1.9	114	12	Q70470	Generic DNA sequence	2.61e-04
c	27	31	1.9	114	12	Q70473	Generic DNA sequence	2.61e-04
	28	31	1.9	1167	6	Q37704	Aphanocladium album m	2.61e-04
	29	31	1.9	1167	11	Q68353	Aphanocladium album m	2.61e-04
	30	31	1.9	1320	6	Q37714	Aphanocladium album m	2.61e-04
	31	31	1.9	1364	6	Q37713	Aphanocladium album m	2.61e-04
	32	31	1.9	1405	6	Q37711	Aphanocladium album p	2.61e-04
	33	31	1.9	1701	6	Q37712	Aphanocladium album p	2.61e-04
	34	30	1.8	1074	12	Q78145	Trichoderma harzianum	8.89e-04
	35	30	1.8	1554	12	Q78144	Trichoderma harzianum	8.89e-04
c	36	28	1.7	114	12	Q70471	Generic DNA sequence	9.80e-03
c	37	27	1.6	39	7	Q51787	Mixed oligonucleotide	3.16e-02
c	38	27	1.6	114	12	Q70472	Generic DNA sequence	3.16e-02
c	39	26	1.6	565	6	Q35072	HCV envelope region n	1.00e-01
	40	27	1.6	3871	2	N71302	HSV-1 gB and surround	3.16e-02
	41	27	1.6	6180	23	T36387	Periplasmic chitodext	3.16e-02
c	42	25	1.5	82	21	T13610	DC43 TSAR library gen	3.11e-01
	43	25	1.5	146	21	T13624	ME#2b library generat	3.11e-01
	44	25	1.5	498	3	N50034	Sequence encoding new	3.11e-01
	45	25	1.5	2951	23	T36390	Soluble chitinase cod	3.11e-01

ALIGNMENTS

RESULT 1
ID Q85245 standard; cDNA; 1681 BP.
AC Q85245;
DT 12-AUG-1995 (first entry)
DE YKL-40 gene.
KW YKL-40; diagnosis; prognosis; therapy; breast cancer; metastasis;
KW marker; joint disease; connective tissue; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 135..1681
FT /*tag= a
PN WO9501995-A.
PD 19-JAN-1995.
PF 08-JUL-1994; U07754.
PR 09-JUL-1993; US-089989.
PA (REGC) UNIV CALIFORNIA.
PI Johansen JS, Price PA;
DR WPI; 95-066866/09.
PT Use of YKL-40 and anti-YKL-40 antibodies - for developing prods.
PT for diagnosis prognosis and therapy of diseases involving
PT connective tissue degradation.
PS Disclosure; Page 66-67; 88pp; English.
CC YKL-40 (40 kDa) was purified from human osteosarcoma MG63 cells.
CC The N-terminal sequence is shown in R70745; the full coding region
CC of the YKL-40 gene is given in Q85245. Homology of the N-terminal
CC and 2 internal peptides (R70746-47) with a bacterial polysaccharide
CC hydrolase suggests that YKL-40 degrades polysaccharide components
CC of connective tissue. YKL-40 is a marker of e.g. metastatic breast
CC cancer and inflammatory or degenerative joint diseases.
SQ Sequence 1681 BP; 401 A; 503 C; 421 G; 356 T;

Query Match 10.6%; Score 176; DB 13; Length 1681;
Best Local Similarity 63.5%; Pred. No. 5.39e-101;
Matches 697; Conservative 0; Mismatches 377; Indels 24; Gaps 10;

```
Db     94 caggctttgtggtcctggtgctgctccagtgctgctctgcatacaaactggtctgctact 153
          |||| ||  |||||||| || || |||  ||  |||||||    ||||||||||||||||
Qy     49 CAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTGCTACT 108

Db    154 acaccagctggtcccagtaccgggaaggcgatgggagctgcttcccagatgcccttgacc 213
          ||||| |||| |||||||| |  | || || |   ||| | | ||  | | | | ||||
```

```
Qy    109 TCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACC 168

Db    214 gcttcctgtgtacccacatcatctacagctttgccaatataagcaacgatcacatcgaca 273

Qy    169 CCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCTGAGCA 228

Db    274 cctgggagtggaatgatgtgacgctctacggcatgctcaacacactcaacaacacgaacc 333

Qy    229 CCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGATGAATC 288

Db    334 ccaacctgaagactctcttgtctgtcggaggatggaactttgggtctcaaagattttcca 393

Qy    289 CCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCAGCACTCAGAAGTTCACAG 348

Db    394 agatagcctccaacacccagagtcgccggactttcatcaagtcagtaccgccatttctgc 453

Qy    349 ATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTTTCTGC 408

Db    454 gcacccatggctttgatgggcgtgaccttgcctggctctaccctgga---c---ggag-- 505

Qy    409 GCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGCCAGGGGAGCC 468

Db    506 -------agacaaacaccatttttaccaccctaatcaaggaaatgaaggccgaatttataa 558

Qy    469 CTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAATGCCTTCCAGC 528

Db    559 aggaagcccag--c-cagggaaaaagcagctcctgctcagcgcagcactgtctgcgggga 615

Qy    529 AGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCCAGCTGGGC 588

Db    616 aggtcaccattgacagcagctatgacattgccaagatatcccaacacctggatttcatta 675

Qy    589 AGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATTTTGTCA 648

Db    676 gcatcatgacctacgattttcatggcgcctggcgtgggaccacaggccatcacagtcccc 735

Qy    649 ACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACATAACAGCCCCC 708

Db    736 tgttccgaggtcaggaggatgcaagtcctgacagattcagcaacactgactatgctgtgg 795

Qy    709 TCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGATGCTGCTGTGC 768

Db    796 ggtacatgttgaggctgggggctcctgccagtaagctggtgatgggcatccccaccttcg 855

Qy    769 AACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATGCCTACCTACG 828

Db    856 ggaggagcttcactctggcttcttc-tgagactg--gtgttccagcgccaatctcaggac 912

Qy    829 GACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCAGCCACAGGGT 888

Db    913 cgggaattccaggccggttcaccaaggaggcagggacccttgcctactatgagatctgtg 972

Qy    889 CTGGCACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTATGAAGTCTGCT 948

Db    973 acttcctccgcggagccacagtccatagaaccctcggccagcaggtcccctatgccacca 1032

Qy    949 CCTGGAAG-G-GG-GCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCCTACATCTTCC 1005

Db   1033 agggcaaccagtgggtaggatacgacgaccaggaaagcgtcaaaagcaaggtgcagtacc 1092

Qy   1006 GGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAGGTCAGCTATC 1065

Db   1093 tgaaggataggcagctggcaggcgccatggtatgggccctggacctggatgacttccagg 1152

Qy   1066 TGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGATGACTTTGCCG 1125

Db   1153 gctccttctgcggccagg 1170

Qy   1126 GCTTCTCCTGCAACCAGG 1143


RESULT    2
ID   N81756 standard; DNA; 966 BP.
AC   N81756;
DT   19-OCT-1990  (first entry)
DE   Gene encoding polypeptide involved in protective mechanisms
KW   Immune response; cell growth; ss.
FH   Key             Location/Qualifiers
FT   CDS             1..966
FT   /*tag= a
PN   J63032898-A.
PD   30-JAN-1988.
PF   16-JUL-1986; 167518.
PR   16-JUL-1986; JP-167518.
PA   (DAIN) Dainippon Pharm KK.
DR   WPI; 88-068419/10.
DR   P-PSDB; P81342.
PT   New polypeptide and DNA encoding it - related to protective mechanisms
PT   such as immune response etc.
PS   Disclosure; p; Japanese.
CC   Polypeptide encoded is involved in protective mechanisms such as immune
CC   response, cell growth and activation of protective functions.
SQ   Sequence   966 BP;   238 A;   282 C;   247 G;   199 T;

  Query Match            7.9%;  Score 131;  DB 1;  Length 966;
  Best Local Similarity  66.1%;  Pred. No. 2.21e-69;
  Matches   269;  Conservative    0;  Mismatches 138;  Indels   0;  Gaps   0;

Db     23 caggctttgtggtcctggtgctgctccagtgctgctctgcatacaaactggtctgctact 82

Qy     49 CAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTGCTACT 108

Db     83 acaccagctggtcccagtaccgggaaggcgatgggagctgcttcccagatgcccttgacc 142

Qy    109 TCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACC 168

Db    143 gcttcctctgtacccacatcatctacagctttgccaatataagcaacgatcacatcgaca 202

Qy    169 CCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCTGAGCA 228

Db    203 cctgggagtggaatgatgtgacgctctacggcatgctcaacacactcaagaacaggaacc 262

Qy    229 CCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGATGAATC 288

Db    263 ccaacctgaagactctcttgtctgtcggaggatggaactttgggtctcaaagattttcca 322

Qy    289 CCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCAGCACTCAGAAGTTCACAG 348

Db    323 agatagcctccaacacccagagtcgccggactttcatcaagtcagtaccgccatttctgc 382

Qy    349 ATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTTTCTGC 408

Db    383 gcacccatggctttgatggggtggaccttgcctggctctaccctgga 429

Qy    409 GCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGA 455


RESULT    3
ID   Q10572 standard; DNA; 1047 BP.
AC   Q10572;
DT   09-APR-1991  (first entry)
DE   Human Natriuretic Peptide Receptor B.
KW   NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW   hyperaldosteronism; glaucoma; guanyl cyclase.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide          1..22
FT   /label= signal sequence
FT   Protein          12
FT   /label= mature NPBR
FT   Domain           23..455
FT   /label= extracellular domain
FT   /note= "binds natriuretic peptides A,B and C]"
FT   Domain           456..456
FT   /label= transmembrane domain
FT   Domain           479..1047
```

```
FT   /label= cytoplasmic domain
FT   /note= "GC and protien kinase activity"
FT   Modified -site  24..26
FT   /label= N-glycos_site
FT   Modified -site  35..37
FT   /label= N-glycos_site
FT   Modified -site  161..163
FT   /label= N-glycos_site
FT   Modified -site  195..197
FT   /label= N-glycos_site
FT   Modified -site  244..246
FT   /label= N-glycos_site
FT   Modified -site  277..279
FT   /label= N-glycos_site
FT   Modified -site  349..351
FT   /label= N-glycos_site
FT   Modified -site  600..602
FT   /label= N-glycos_site
PN   WO9100292-A.
PD   10-JAN-1991.
PF   22-JUN-1990; U03586.
PR   23-JUN-1989; US-370673.
PA   (GETH ) GENENTECH INC.
PI   Chang M, Goeddel D, Lowe D;
DR   WPI; 91-036711/05.
DR   N-PSDB; Q10324.
PT   Natriuretic protein receptor B - for diagnosis and treatment of
PT   kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS   Claim 3; Fig 1; 49pp; English.
CC   The sequence was derived from the DNA encoding natriuretic peptide
CC   receptor B, NPRB, having guanyl cyclase (GC) activity and protein
CC   kinase activity.  The DNA can be inserted into expression vectors
CC   for the prodn. of the protein, opt. after being mutated to produce
CC   NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC   114,952).   The protein (or variants) can be used in treatment of
CC   natriuretic peptide disorders, and also to isolate peptides using
CC   affinity chromatography.  Antibodies with affinity for NPRB can
CC   also be prepd.
SQ   Sequence   1047 BP;    87 A;    15 C;    83 G;    51 T;

  Query Match            5.6%;  Score 92;  DB 2;  Length 1047;
  Best Local Similarity  10.6%;  Pred. No. 1.24e-42;
  Matches   106;  Conservative  279;  Mismatches 603;  Indels  15;  Gaps  14;

Db      5 snnnnvaanaggvrnngarnntnavvnnnhnnsyawawnrvgnavanavnangrannvdn 64
Qy    120 GCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACCCCAGCCTTTGC 179

Db     65 rnvssnnngacsnynannsavdnknyhdndnnngngcvynaas-varnashwrnnnntag 123
Qy    180 ACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCTGAGCACCACTGAGTGG 239

Db    124 avasgnsaknd hyrtnvrtgnsanknqnnvvtnhghnnwtaraannyndartddrnhynt 183
Qy    240 AATGACGAG-ACTCT-C-TACCAGGAGTTCAATGGCCTGAAGAAGATGAATCCCAAGCTG 296

Db    184 nngvnnanngsnnsvnhnvyarnnggnnnathnnrangrnvyncgnnnmnhnnnnnan-r 242
Qy    297 AAGACCCTGTTAGCCATCGGAGGCTGGAATTTCAGCACTCAGAAGTTCACAGATATGGTA 356

Db    243 nnntngdyvnnyndvngnsnragntratgrnwndnrtrnnananrnanntvnvntyrnnn 302
Qy    357 GCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTTTCTGCGCAAATAC 416

Db    303 nnnynnnnnrnnnrarndngvnngnsnmnnnagcnydgnnnyanvnnntnnnggtrndgn 362
Qy    417 AGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGCCAGGGGAGCCC-TGCCGT 475

Db    363 rnvnkmngrryhgvtgnvvmdknndrntdnvnwamgdndsgdnnnaahysganknnwwtg 422
Qy    476 AGACAAGGAGC-GCTTCACAACCCTGGTACAGGACTTGGCCAATGCCTTCCAGCAGGAAG 534

Db    423 rnnn-wv-kgannsdnnncandnddnscdktnnstnanvangtgntnnmngvssnnnnrk 480
Qy    535 CCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCCAGCTGGGCAGACCT 594

Db    481 nmnnknnasmnwrnrwnnnnngnsnryhkgagsrntnsnrgssygsnmtahgkynnnant 540
Qy    595 ATGTGGATGC-TGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATTTTGTCAACCTT 653

Db    541 ghnkgnvvankhvnkkrnnntrnvnnnnkhmrdvnnnhntrnngacndnnnncnvtnycn 600
Qy    654 -ATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACATAACAGCCCCCTCTA 712

Db    601 rgsnndnnnndsnnndwmnrysnnndnvkgmannhnsnnsshgsnkssncvvdsrnvnkn 660
Qy    713 CAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGATGCTGCTGTGCAACA 772

Db    661 tdygnasnrstannddnnanyakknntannnnsgnnnnttgmnaadvysngnnnnnnanr 720
Qy    773 GTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATGCCTACCTACGGACG 832

Db    721 sgnnynngndnsnknnvnkvrngnrnynrnsndrtnnnnnnvnnmnrcwandnanrndng 780
Qy    833 CTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCAGCCACAGGGTCTGG 892

Db    781 nnkgnnrrnnknggtsnndnnnnrmnnyannnnknvnnrtnaynnnkrkanannynnnnh 840
Qy    893 CACTCCAGGCCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTATGAAGTCTGCTCCTG 952

Db    841 svannnkrgntvnanandsvtnynsdnvgntansanstnmnvvtnnndnytcnd--annd 898
Qy    953 GAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCCTACATCTTCCGGGACAA 1012

Db    899 nndvykvntngdaymvvsgnngrngnrhannnarmananndavssnrnrhrnhdnnrnrn 958
Qy   1013 CCAGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAGGTCAGCTATCTGAAGCA 1072

Db    959 gvhtgnvcagvvgnkmnrycnngdtvntasrmnsngnanknhv 1001
Qy   1073 GAAGGGACTGGGCGG-GGCCATGGTC-TGGGCACTGGACTTAG 1113


RESULT    4
ID   Q90443 standard; cDNA to mRNA; 2504 BP.
AC   Q90443;
DT   10-JAN-1996  (first entry)
DE   Murine oviduct specific glycoprotein cDNA.
KW   Murine oviduct specific glycoprotein; recombinant production;
KW   MOGP; ds.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   CDS             3..2160
FT   /*tag=  a
FT   /note= "START codon absent"
FT   sig_peptide     3..56
FT   /*tag=  b
FT   mat_peptide     57..2156
FT   /*tag=  c
PN   J07107979-A.
PD   25-APR-1995.
PF   15-AUG-1994; 214227.
PR   19-AUG-1993; JP-227881.
PA   (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR   WPI; 95-190179/25.
DR   P-PSDB; R90443.
PT   New DNA encoding an oviduct-specific glycoprotein - useful for
PT   recombinant protein production in high quantities.
PS   Claim 1; Pages 11-14; 22pp; Japanese.
CC   Q90443 encodes R73992 murine oviduct specific glycoprotein (MOGP).
CC   The cDNA can be used for the commercial recombinant prodn. of
CC   MOGP in high quantities.
SQ   Sequence   2504 BP;    678 A;    616 C;    538 G;    672 T;

  Query Match            5.5%;  Score 91;  DB 15;  Length 2504;
  Best Local Similarity  58.5%;  Pred. No. 5.83e-42;
```

```
  Matches   363;  Conservative    0;  Mismatches 254;  Indels   3;  Gaps   2;

Db      60 aaactggtgtgctatttcaccaattgggcacatagtcggccaggccctgcctccatcatg 119
           |||||||| ||||| |||||||| ||||| ||      | |  ||    ||   | || ||
Qy      93 AAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTG 152

Db     120 ccccatgacctggaccccttttctttgtacgcatctgatatttgcctttgcctcaatgagc 179
           ||| | ||| |||||||||   ||||| || || || || |  ||||| ||    |||| |
Qy     153 CCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACC 212

Db     180 aacaatcagattgtggccaagaatctgcaggatgaaaacgttctctatccagagttcaac 239
           ||| | ||| | | |  ||  | |  |  | ||||  |   ||||||| |  ||||||||
Qy     213 AACCACCAGCT-GAG--CACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAAT 269

Db     240 aaactcaaggaaaggaacagggagctgaaaacactgctgtccattggagggtggaacttc 299
              || ||| | | |||     |||||||| || ||| |  |||| |||| ||||| |||
Qy     270 GGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTC 329

Db     300 ggcacatcacggttcaccgctatgctctccacacttgccaaccgggaaaagtttattgac 359
            ||||      |||||| | |||| |  ||||      ||||||  | |  ||| |  ||
Qy     330 AGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAAC 389

Db     360 tcagtcatatccttcctgagaatacatggttttgatggtcttgacctttttcttcttgtac 419
           || | |||    || ||| | | | |  | ||||| || |||||||||  ||    ||||
Qy     390 TCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTAC 449

Db     420 cctggactacgaggcagtcccccacacgaccggtggaatttttctcttcttaattgaagag 479
           || |||   |  || ||  |   |||  |  |   ||      | |  |  | ||
Qy     450 CCAGGAAGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGAC 509

Db     480 ctccagtttgccttcgagagggaggccctgctcactcagcacccacggctgctgctgtcg 539
           |     |||||||  ||  ||| ||||  |  | |   | |   ||| || || |||
Qy     510 TTGGCCAATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGT 569

Db     540 gctgctgtctctggcatcccaagcatcatccatacgtcttatgatgcgctccttttagga 599
           || || ||  | |     |  | |     |  || |    || || | |  |    | |
Qy     570 GCAGCGGTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCC 629

Db     600 aggcgtttggatttcattaatgttttgtcttatgacttacatggaagttgggaaaagttt 659
             |   |||||||  | ||  || || | || ||||| |||||   ||||||| ||| |
Qy     630 CAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTC 689

Db     660 acaggacacaatagtcctct 679
           || ||||| || || || ||
Qy     690 ACGGGACATAACAGCCCCCT 709


RESULT    5
ID   Q90442 standard; cDNA to mRNA; 1994 BP.
AC   Q90442;
DT   10-JAN-1996  (first entry)
DE   Bovine oviduct specific glycoprotein cDNA.
KW   Bovine oviduct specific glycoprotein; recombinant production;
KW   BOGP; ds.
OS   Bos taurus.
FH   Key             Location/Qualifiers
FT   CDS             2..1615
FT   /*tag=  a
FT   /note= "START codon absent"
FT   sig_peptide     2..55
FT   /*tag=  b
FT   mat_peptide     56..1612
FT   /*tag=  c
PN   J07107979-A.
PD   25-APR-1995.
PF   15-AUG-1994; 214227.
PR   19-AUG-1993; JP-227881.
PA   (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR   WPI; 95-190179/25.
DR   P-PSDB; R73991.
PT   New DNA encoding an oviduct-specific glycoprotein - useful for
PT   recombinant protein production in high quantities.
PS   Claim 1; Pages 7-9; 22pp; Japanese.

CC   Q90442 encodes R73991 bovine oviduct specific glycoprotein (BOGP).
CC   The cDNA can be used for the commercial recombinant prodn. of
CC   BOGP in high quantities.
SQ   Sequence   1994 BP;    445 A;    536 C;    520 G;    493 T;

  Query Match             5.0%;  Score 83;  DB 15;  Length 1994;
  Best Local Similarity  57.9%;  Pred. No. 1.33e-36;
  Matches   359;  Conservative    0;  Mismatches 258;  Indels   3;  Gaps   2;

Db      59 aagctggtgtgttatttcaccaactgggcattcagtcggcccggccctgcctcaatcctg 118
           || ||||| || || ||||||||||||||        | |  ||    ||     |||||
Qy      93 AAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTG 152

Db     119 cctcgggatctggacccgtttctctgcacccacctggtatttgcctttgcctcgatgagc 178
           ||   |||  |||||||    || ||||||||||  | |  ||||| ||    |||| |
Qy     153 CCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACC 212

Db     179 aacaatcagattgttcctaaggatccccaggatgagaaaatcctctacccagagttcaac 238
           ||| | ||| |     | |  ||     | || ||||   || ||||||  ||||||||
Qy     213 AACCACCAGCTGAGCACCACTGAGTGGAATGACGAGAC--TC-TCTACCAGGAGTTCAAT 269

Db     239 aagctcaaggagaggaacagagggctgaaaacgctgctgtccatcggggggtggaacttt 298
           || ||| ||| |||      |||||| || ||| |  ||||||| || ||||| ||
Qy     270 GGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTC 329

Db     299 ggcaccgtgaggttcaccacgatgctgtccacgttttccaaccgggaaaggttcgtcagt 358
           ||||   || ||||||    ||| |  |||||      ||||||  | |  || ||||
Qy     330 AGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAAC 389

Db     359 tcagtgatcgccctcctgaggacacatggctttgatggtctggacctcttcttcttgtac 418
           || |  |||    | ||| | | | |  ||||||| || || |||||   ||    ||||
Qy     390 TCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTAC 449

Db     419 cctggactcagaggcagccctgcgcgtgaccgctggacctttgtcttcttacttgaagag 478
           || |||  |   || |||||||||    |||      |  |||      | |  |  | ||
Qy     450 CCAGGAAGCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGAC 509

Db     479 ctcctgcaggccttcaagaatgaggcccagctcaccatgcgtccaaggctgctgctgtct 538
           |     | |||||| || | || ||||||  | |   |     | | || || |||  |
Qy     510 TTGGCCAATGCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGT 569

Db     539 gctgccgtctctggggacccccatgtcgtccagaaagcgtatgaagcacgccttctgggc 598
           || || ||  | |  |  |         || |    |  || || |    |    | | |
Qy     570 GCAGCGGTTCCAGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCC 629

Db     599 agactcctggatttcatcagtgtcttgtcttatgacttacatggaagctgggaaaaggtc 658
                 |||||||||  |||   |  || | || ||||| |||||    ||||| ||||||
Qy     630 CAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTC 689

Db     659 acaggacacaatagccctct 678
           || ||||| || ||||| ||
Qy     690 ACGGGACATAACAGCCCCCT 709


RESULT    6
ID   Q10572 standard; DNA; 1047 BP.
AC   Q10572;
DT   09-APR-1991  (first entry)
DE   Human Natriuretic Peptide Receptor B.
KW   NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW   hyperaldosteronism; glaucoma; guanyl cyclase.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT   /label= signal sequence
FT   Protein         12
FT   /label= mature NPBR
FT   Domain          23..455
FT   /label= extracellular domain
FT   /note= "binds natriuretic peptides A,B and C]"
FT   Domain          456..456
FT   /label= transmembrane domain
FT   Domain          479..1047
```

```
FT   /label= cytoplasmic domain
FT   /note= "GC and protien kinase activity"
FT   Modified -site  24..26
FT   /label= N-glycos_site
FT   Modified -site  35..37
FT   /label= N-glycos_site
FT   Modified -site  161..163
FT   /label= N-glycos_site
FT   Modified -site  195..197
FT   /label= N-glycos_site
FT   Modified -site  244..246
FT   /label= N-glycos_site
FT   Modified -site  277..279
FT   /label= N-glycos_site
FT   Modified -site  349..351
FT   /label= N-glycos_site
FT   Modified -site  600..602
FT   /label= N-glycos_site
PN   WO9100292-A.
PD   10-JAN-1991.
PF   22-JUN-1990; U03586.
PR   23-JUN-1989; US-370673.
PA   (GETH ) GENENTECH INC.
PI   Chang M, Goeddel D, Lowe D;
DR   WPI; 91-036711/05.
DR   N-PSDB; Q10324.
PT   Natriuretic protein receptor B - for diagnosis and treatment of
PT   kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS   Claim 3; Fig 1; 49pp; English.
CC   The sequence was derived from the DNA encoding natriuretic peptide
CC   receptor B, NPRB, having guanyl cyclase (GC) activity and protein
CC   kinase activity.  The DNA can be inserted into expression vectors
CC   for the prodn. of the protein, opt. after being mutated to produce
CC   NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr-
CC   114,952).   The protein (or variants) can be used in treatment of
CC   natriuretic peptide disorders, and also to isolate peptides using
CC   affinity chromatography.  Antibodies with affinity for NPRB can
CC   also be prepd.
SQ   Sequence   1047 BP;    87 A;    15 C;    83 G;    51 T;

  Query Match            4.7%;  Score 78;  DB 2;  Length 1047;
  Best Local Similarity   9.5%;  Pred. No. 2.76e-33;
  Matches    91;  Conservative  267;  Mismatches 590;  Indels  12;  Gaps  11;

Db     86 dnknyhdndnnngngcvynaasvarnashwrnnnntagavasgnsakndhyrtnv-rtgn 144
          : : ::: :   | ||:: ||::|:   ::     |  |:  |  |: : ::  : : |
Cp   1655 TTTTTTTTTGAGGGGCACAAACCAAAGATTTATTTTGCAAGTGAAAGGGGCAGCCCAGGA 1596

Db    145 sankngnnvvtnhghnnwtaraannyndartddrnhyntnngvnnanngsnnsvnhnvya 204
          :|      ::  :|     :|  : : :  :: :     |:  |  |:  :: : :
Cp   1595 GACCCAGAAAAAAGGAAGGCAAGGCTGAGAGCAGAAAGCCTGGATAAAGGAAGACCACAG 1536

Db    205 rnnggnnnathnnrangrnvyncgnnnmnhnnnnnanrnnntngdyvnnyndvngn-snr 263
          :  ||     :  :   : ::      : :     | :     |: :    :: |  : :
Cp   1535 AAAGGCCTGCAGGAGCCAGATTGCGGCCCCCAGGGAAAACCCAGGAGGCAGGCTGTAGAG 1476

Db    264 agntratgrnwndnrtrnnananrnanntvnvntyrnnnnnnynnnnnrnnnrarndngv 323
          | |  ||:     :    | |       : :   :       :          :     |:
Cp   1475 TGATCCTGGGCCCAGCCTCAAAGCTGGGACTGGA-GGGGCTTTAGCGACTCAATTCCAGG 1417

Db    324 nngnsnmnnnagcnydgnnnyanvnnntnnnggtrndgnrnvnkmngrryhgvtgnvvmd 383
          : :      :|   :  :       |    :  : : ::   :::|:   :::
Cp   1416 TGCAGCATTTGCAGGAGTTGCTGAACACCAGGCCTGTCGGGCAGCTTTGCTGGAACAGCC 1357

Db    384 knnd-rntdnvnwamgdndsgdnnnaahysganknnwwtgrnnnwvkgannsdnnncand 442
          :    : |: : : : : :: :     :::|      : |:    :  |  :       :
Cp   1356 GCCCCGCTGCACAGCTGTAGAAGCTGGACCGTTCCCGAGGATTGGGATAGAGCCCATCAG 1297

Db    443 nddnscdktnnstnanvangtgntnnmngvssnnnnrknmnnknn-asmnwrnrwnnnnn 501
          :: :|  |  :  | :  ||  |  : :::    :: : :   |:  :  ::
Cp   1296 CTTTGCCCTGGCAGAACGTGTCTTGTCCAGGGCTGGGGCCATGCTCAGGTTCAGAGGGCT 1237

Db    502 gnsn--ryhkgagsrntnsnrgs-sygsnmtahgkynnnant-ghnkgnvvankhvnkkr 557
```

```
          | :   ::::  | : | : :|: ::|:  | : :        |: :| ::   :: ::
Cp   1236 GACCTGGTTTTGGAACTTCAAGCTCTGGGGTGCCTGAAGGCAAGTATGGAAGACTCAGTT 1177

Db    558 nnntrnvnnnnkhmrdvnnnhntrnngacndnnnncnvtnycnrgsnndnnnndsnnndw 617
          :    ::: ::   :  :   |  :      :|  | :|   :    ::   ::
Cp   1176 CCTGCCGTAGCGTCTGGATGAGGGGGTATCGGCCCTGGTTGCAGGAGAAGCCGGCAAAGT 1117

Db    618 mnrysnnndnvkgmannhnsnnsshgsnkssncvvdsrnvnkntdygnasnrstannddn 677
          :  :    : : |    : :  : :|:  :: |::::  : :  :    : : ||  ::
Cp   1116 CATCTAAGTCCAGTGCCCAGACCATGGCCCCGCCCAGTCCCTTCTGCTTCAGATAGCTGA 1057

Db    678 nanyakknnt-annnnsgnnnnttgmn-aadvysngnnnnnnnanrsgnnynngndnsnkn 735
          : ::  | |    :      | :  ||::::           |  :  | :   :
Cp   1056 CCTTGGTTTTGAAGCTCTCCACATCATCAAAGCCCACCCACTGGTTGTCCCGGAAGATGT 997

Db    736 nvnkvrngnrnynrnsndrtnnnnnnvnnmnrcwandnanrndngnnkgnnrrnnknggt 795
          : :::    : : : ::          : :      :   :    ::  :   |
Cp    996 AGGGCACCTTCTGATCCTGGATTCTCTGTTTGGTGGCCCCCTTCCAGGAGCAGACTTCAT 937

Db    796 snndnnnnrmnnyannnnknvnnrtnaynnnkrkanannynnnnhsvannnkrgntvnan 855
          :    :  :              |  :  :::          :::      |  :
Cp    936 AGTAGGCCAGCATCCCTCCTTCCTTGGTGAAGGGGCCTGGAGTGCCAGACCCTGTGGCTG 877

Db    856 andsvtnynsdnvgntansanstnmnvvtnnndnytcndanndnndvykvntngdaymvv 915
            :::  : ::  | |  :   | : ::    : :   :|  :  : : :   |:|  ::
Cp    876 GGGCCCCCACTCTGGTGTCTGATGAGGAGGCCAGTGTGAAGGAGCGTCCGTAGGTAGGCA 817

Db    916 sgnngrngnrhannnarmananndavssnrnrhrnhdnnrnrngvhtgnvcagvvgnkmn 975
          |   : |  :|    ::     : :::    :  ::  :    ::   :|| ::  ::
Cp    816 TGCCAAGGATCAGCTTGCTGGCAG-GGGTCCCCTTCTGCAGCCACTGTTGCACAGCAGCA 758

Db    976 rycnngdtvntasrmnsngnanknhvssttkdandnngcnnnnnrgdvnmkgkgkmrtyw 1035
          :|  |:|:   :   : |     :: :| :      ||     : ::  :|:|  :|::
Cp    757 TCCACGTTGAGGCTGGCTGCTGCACCACTCTCTTCTTGCCTCTTGTAGAGGGGGCTGTTA 698
```

```
RESULT    7
ID   Q90444 standard; cDNA to mRNA; 2366 BP.
AC   Q90444;
DT   10-JAN-1996  (first entry)
DE   Hamster oviduct specific glycoprotein cDNA.
KW   Hamster oviduct specific glycoprotein; recombinant production;
KW   HOGP; ds.
OS   Cricetulus griseus.
FH   Key             Location/Qualifiers
FT   CDS             3..2009
FT   /*tag= a
FT   /note= "START codon absent"
FT   sig_peptide     3..56
FT   /*tag= b
FT   mat_peptide     57..2006
FT   /*tag= c
PN   J07107979-A.
PD   25-APR-1995.
PF   15-AUG-1994; 214227.
PR   19-AUG-1993; JP-227881.
PA   (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR   WPI; 95-190179/25.
DR   P-PSDB; R73993.
PT   New DNA encoding an oviduct-specific glycoprotein - useful for
PT   recombinant protein production in high quantities.
PS   Claim 1; Pages 16-18; 22pp; Japanese.
CC   Q90444 encodes R73993 hamster oviduct specific glycoprotein (HOGP).
CC   The cDNA can be used for the commercial recombinant prodn. of
CC   HOGP in high quantities.
SQ   Sequence   2366 BP;    602 A;    574 C;    573 G;    617 T;

  Query Match            4.0%;  Score 67;  DB 15;  Length 2366;
  Best Local Similarity  60.8%;  Pred. No. 4.37e-26;
  Matches   231;  Conservative    0;  Mismatches 146;  Indels   3;  Gaps   2;

Db     60 aagctggtctgctatttcaccaactgggctcacagtcggccagtccctgcctccatcttg 119
          || |||||||||| |||||||||||||| ||     | |  |     ||   | || ||
```

```
Qy      93 AAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTG 152

Db     120 ccccgtgacctggatccctttctttgtacacacctgatatttgcctttgcctcgatgagc 179
           |||   ||| |||| |||  ||||| || ||||| || |  ||||| ||    |||| |
Qy     153 CCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACC 212

Db     180 aacaatcagattgttgccaataatctccaggatgagaaaattctctatccagagttcaac 239
           ||| | ||| |     ||| | |  |  |  |||| | | |||||| |  ||||||||
Qy     213 AACCACCAGCTGAGCACCACTGAG-TGGA--ATGACGAGACTCTCTACCAGGAGTTCAAT 269

Db     240 aaactcaaggagaggaacagagccctgaaaacactactgtctgttggaggctggaacttc 299
             || ||| ||| |||       ||||| || ||  |  |  | ||||||||||| |||
Qy     270 GGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTC 329

Db     300 ggcacatcacggttcaccactatgctgtccacccttgccagccgtgaaaaatttattggt 359
           ||||      ||||||   |||| |  ||||     || |||| | |  ||| |
Qy     330 AGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAAC 389

Db     360 tcagttgtatccttcctgagaacacatggctttgatgggcttgatctcttcttcttgtac 419
           || |   |    || ||| | | | |  ||||||| || ||||| ||   ||    ||||
Qy     390 TCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTAC 449

Db     420 cctggactacgaggcagccc 439
           || |||   |  || |||||
Qy     450 CCAGGAAGCCAGGGGAGCCC 469


RESULT    8
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT   /*tag= a
FT   /function=multiple cloning site
FT   primer_bind     187..204
FT   /*tag= b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match            3.1%;  Score 51;  DB 1;  Length 204;
  Best Local Similarity  15.4%;  Pred. No. 5.90e-16;
  Matches    18;  Conservative   62;  Mismatches  36;  Indels   1;  Gaps   1;

Db     75 yycdchvgccgymrttthhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhdd 134
          ::|: :: | | : ||| ::::: : :::::     | :::|  ::::::    | ::::
Qy    628 CCCAGAACCTGGATTTTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGG 687

Db    135 hyvybbbvynvhnhnncncccbnnhvchnvhbnnhrnwayvrhdarrddvhccvchc 191
          ::::::::  : :  | |||:  ::|: :::  :: :| :::: :::::: |: :|
Qy    688 TCACGGGACATAACAGCCCCCTC-TACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCC 743


RESULT    9
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match            2.7%;  Score 45;  DB 9;  Length 91;
  Best Local Similarity  2.0%;  Pred. No. 2.64e-12;
  Matches     1;  Conservative   47;  Mismatches   3;  Indels   0;  Gaps   0;

Db     10 gssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsv 60
          |:::::::::: :::::: :::::::::::::::::::::::::::: :::::::
Cp   1386 GGCCTGTCGGGCAGCTTTGCTGGAACAGCCGCCCCGCTGCACAGCTGTAGA 1336


RESULT   10
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match            2.5%;  Score 42;  DB 9;  Length 91;
  Best Local Similarity  3.8%;  Pred. No. 1.60e-10;
  Matches     2;  Conservative   45;  Mismatches   5;  Indels   0;  Gaps   0;

Db     12 svhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsvctc 63
          :::::::::: :::::::::::::   :::::::::::::::::::: ||
Qy    149 CCTGCCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTC 200
```

```
RESULT   11
ID   Q70468 standard; DNA; 114 BP.
AC   Q70468;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT   /*tag= a
FT   /note= "this sequence represents 'Z'; Z can be a
FT   sequence of 6, 9 or 12 nucleotides (see
FT   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65154.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed activity
CC   allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match             2.3%;  Score 38;  DB 12;  Length 114;
  Best Local Similarity   4.5%;  Pred. No. 3.39e-08;
  Matches     5;  Conservative   34;  Mismatches  73;  Indels   0;  Gaps   0;

Db       3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnn 62
           :  :  :  :  :  :  :  :  :  :  :|||  :  :  :  :  :  :
Qy     146 CTTCCTGCCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGG 205

Db      63 bnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
           :  :  :  :  :  :  : ||  :  :  :  :  :  :  :  :  :  :  :
Qy     206 CATGACCAACCACCAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTAC 257


RESULT   12
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT   /*tag= a
FT   /function=multiple cloning site
FT   primer_bind     187..204
FT   /*tag= b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match             2.3%;  Score 38;  DB 1;  Length 204;
  Best Local Similarity  11.5%;  Pred. No. 3.39e-08;
  Matches    11;  Conservative   49;  Mismatches  36;  Indels   0;  Gaps   0;

Db      91 thhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddhyvybbbvynvhnhnn 150
           |:::::: : :  :  :::| |: || :  :  :| :   : : : ::::::  :: :
Cp    1150 TATCGGCCCTGGTTGCAGGAGAAGCCGGCAAAGTCATCTAAGTCCAGTGCCCAGACCATG 1091

Db     151 cncccbnnhvchnvhbnnhrnwayvrhdarrddvhc 186
             |||:  :: : :::  :: : ::::::|: ::::|
Cp    1090 GCCCCGCCCAGTCCCTTCTGCTTCAGATAGCTGACC 1055


RESULT   13
ID   Q70467 standard; DNA; 114 BP.
AC   Q70467;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT   /*tag= a
FT   /note= "this sequence represents 'Z'; Z can be a
FT   sequence of 6, 9 or 12 nucleotides (see
FT   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65153.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
```

```
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)16(TGC)(NNB)1Z(NNB)16(TGC)(NNB)1Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need for
CC   complex methods of hybridoma formation or in vivo antibody production.
CC   The TSARs are easily characterised and have designed activity allowing
CC   direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match             2.2%;  Score 37;  DB 12;  Length 114;
  Best Local Similarity   4.8%;  Pred. No. 1.26e-07;
  Matches     5;  Conservative   32;  Mismatches  68;  Indels   0;  Gaps   0;

Db      6 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnnnnnnnnbnn 65
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|||  :        :
Qy    134 GGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTA 193

Db     66 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtg 110
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :||
Qy    194 CGCCTTCGCTGGCATGACCAACCACCAGCTGAGCACCACTGAGTG 238


RESULT   14
ID   Q70469 standard; DNA; 114 BP.
AC   Q70469;
DT   07-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT   /*tag=  a
FT   /note= "this sequence represents 'Z'; Z can be a
FT   sequence of 6,9 or 12 nucleotides (see
FT   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC   This generic formula can be represented  as follows: X(TGC)(NNB)10-
CC   (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking  restriction
CC   sites (X is not the same as Y) that are not specified further. This
CC   sequence generates peptides that are cloverleaf in structure. Other
CC   generic sequences are shown in Q70465-68. Other specific peptides
CC   generated by these generic sequences are shown in R65150-54. TSARs are
CC   concatenated heterofunctional proteins or peptides, comprising at least
CC   two functional regions - a binding domain with affinity for a ligand and
CC   a second effector peptide portion that is chemically or biologically
CC   active.They may further comprise a linker peptide between the 2 domains.
CC   The oligonucleotides are also designed so that the expressed peptide
CC   contains 2 or 4 cysteine residues positioned in, or flanking, the
CC   unpredicted or variant residues. These residues confer some degree of
CC   conformational rigidity to the peptides. The TSARs or compsns. comprising
CC   a TSAR binding domain can be used in vivo to deliver a chemically or
CC   biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC   or enzyme, to the specific target or on the cell. They can also replace
CC   the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC   and therefore circumvent the need for complex methods of hybridoma
CC   formation or in vivo antibody production. The TSARs are easily
CC   characterised and have designed activity allowing direct and rapid
CC   detection in a screening process.
SQ   Sequence   114 BP;    0 A;    4 C;    4 G;    4 T;

  Query Match             2.2%;  Score 37;  DB 12;  Length 114;
  Best Local Similarity   6.3%;  Pred. No. 1.26e-07;
  Matches     7;  Conservative   32;  Mismatches  72;  Indels   0;  Gaps   0;

Db      2 gcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnn 61
            ||  :  :  :  :  :  :  :  :  :  :|||  :  :  :  :  :  :
Qy    145 GCTTCCTGCCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTG 204

Db     62 nbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbt 112
          :  :  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :|
Qy    205 GCATGACCAACCACCAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCT 255


RESULT   15
ID   Q70465 standard; DNA; 114 BP.
AC   Q70465;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT   /*tag=  a
FT   /note= "this sequence represents 'Z'; Z can be a
FT   sequence of 6, 9 or 12 nucleotides (see
FT   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65150 and R65151.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
```

CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compsns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARs are easily characterised and have designed
CC activity allowing direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 2.2%; Score 36; DB 12; Length 114;
Best Local Similarity 3.6%; Pred. No. 4.64e-07;
Matches 4; Conservative 34; Mismatches 74; Indels 0; Gaps 0;

```
Db      3 bnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :  :  :  :  : ||  :  :  :  :  :  :  :  :  :  :  :  :
Qy    155 CAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAA 214

Db     63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnb 114
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :| |  :  :  :
Qy    215 CCACCAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTC 266
```

Search completed: Sun Jun 29 20:48:37 1997
Job time : 185 secs.

****

MPSRCH (TM)

****

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Jun 29 20:49:03 1997; MasPar time 1172.65 Seconds
904.203 Million cell updates/sec

Tabular output not generated.

Title: >US-08-663-618A-3
Description: (1-1656) from US08663618A.seq
Perfect Score: 1656
N.A. Sequence: 1 GCTGCAGCCTGCCGCTGAGC.....TGTGCCCCTCAAAAAAAAAA 1656
Comp: CGACGTCGGACGGCGACTCG.....ACACGGGGAGTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 886179 seqs, 320143548 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: EST-STS
1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98
99:EST99
Database: EST-STS-TWO
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134
135:EST135 136:EST136 137:EST137 138:EST138 139:EST139
140:EST140 141:EST141 142:EST142 143:EST143 144:EST144
145:EST145 146:EST146 147:EST147 148:EST148 149:EST149
150:EST150 151:EST151 152:EST152 153:EST153 154:EST154
155:EST155 156:EST156 157:EST157 158:EST158 159:EST159
160:EST160 161:EST161 162:EST162 163:EST163 164:EST164
165:EST165 166:EST166 167:EST167 168:EST168 169:EST169
170:EST170 171:EST171 172:EST172 173:EST173 174:EST174
175:EST175 176:EST176 177:EST177 178:EST178 179:EST179
180:EST180 181:EST181 182:EST182 183:EST183 184:EST184
185:EST185 186:EST186 187:EST187 188:EST188 189:EST189
190:EST190 191:EST191 192:EST192 193:EST193
Database: EST-STS-THREE
194:STS1 195:STS2 196:STS3 197:STS4 198:STS5 199:STS6
200:STS7 201:STS8 202:STS9 203:STS10 204:STS11 205:STS12
206:gnEST1 207:gnEST2 208:gnEST3 209:gnEST4 210:gnEST5
211:gnEST6 212:gnEST7 213:gnEST8 214:gnEST9 215:gnEST10
216:gnEST11 217:gnEST12 218:gnEST13 219:gnEST14
220:gnEST15 221:gnEST16 222:gnSTS 223:enEST1 224:enEST2
225:enEST3 226:enEST4 227:enEST5 228:enEST6 229:enEST7
230:enEST8 231:enEST9 232:enEST10 233:enEST11
234:enEST12 235:enEST13 236:enEST14 237:enEST15
238:enEST16 239:enEST17 240:enEST18 241:enEST19
242:enEST20 243:enEST21 244:enEST22 245:enEST23
246:enEST24 247:enEST25 248:enEST26 249:enEST27
250:enSTS

Statistics: Mean 11.532; Variance 2.180; scale 5.290

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
c	1	345	20.8	428	151	T94579	ye31g05.s1 Homo sapie	0.00e+00
	2	245	14.8	405	151	T94272	ye31g05.r1 Homo sapie	0.00e+00
c	3	215	13.0	241	250	HSG124	human STS SHGC-17364.	0.00e+00
c	4	215	13.0	241	202	G29124	human STS SHGC-17364.	0.00e+00
	5	79	4.8	405	41	H43270	yp05a12.r1 Homo sapie	1.44e-100
	6	78	4.7	541	113	N75653	yv29c01.r1 Homo sapie	1.34e-98
	7	78	4.7	541	207	N75653	yv29c01.r1 Homo sapie	1.34e-98
c	8	73	4.4	533	144	T65854	yc24b03.s1 Homo sapie	7.80e-89
c	9	73	4.4	533	200	G24226	human STS WI-14165.	7.80e-89
	10	71	4.3	311	79	N40107	yw72b09.r1 Homo sapie	5.83e-85
	11	71	4.3	336	169	W51302	ma16f08.r1 Life Tech	5.83e-85
	12	66	4.0	266	35	H24821	yl42f01.r1 Homo sapie	2.27e-75
	13	65	3.9	396	134	R55530	yj79c12.r1 Homo sapie	1.81e-73
	14	60	3.6	306	215	W10705	ma47g07.r1 Soares mou	4.65e-64
	15	60	3.6	306	174	W10705	ma47g07.r1 Soares mou	4.65e-64
	16	60	3.6	421	139	R74158	yj99g05.r1 Homo sapie	4.65e-64
c	17	60	3.6	438	76	N26985	yw72b09.s1 Homo sapie	4.65e-64
	18	60	3.6	586	170	W53184	md19a03.r1 Soares mou	4.65e-64
	19	59	3.6	589	193	W97818	mg03c04.r1 Soares mou	3.38e-62
	20	56	3.4	320	144	T66009	yc24b03.r1 Homo sapie	1.17e-56
	21	49	3.0	299	151	T91693	ye25h05.r1 Homo sapie	4.86e-44
c	22	43	2.6	459	31	H10989	ym07a06.s1 Homo sapie	1.22e-33
	23	41	2.5	461	56	H94102	yw58g11.r1 Homo sapie	2.80e-30
c	24	38	2.3	392	95	N64805	yz31c01.s1 Homo sapie	2.36e-25
	25	36	2.2	335	47	CELK104H2F	C.elegans cDNA clone	3.71e-22
	26	35	2.1	280	98	R03212	pk05a03.s1 Caenorhabd	1.38e-20
	27	35	2.1	360	46	CELK084F3F	C.elegans cDNA clone	1.38e-20
	28	34	2.1	360	47	CELK104C7F	C.elegans cDNA clone	4.86e-19
	29	34	2.1	360	45	CELK051H2F	C.elegans cDNA clone	4.86e-19
	30	34	2.1	429	44	CELK009D7F	C.elegans cDNA clone	4.86e-19
	31	33	2.0	360	45	CELK060G3F	C.elegans cDNA clone	1.63e-17
	32	33	2.0	360	47	CELK109D2F	C.elegans cDNA clone	1.63e-17
	33	32	1.9	323	16	AA067770	mm24c01.r1 Stratagene	5.19e-16
	34	32	1.9	323	244	MMA67770	mm24c01.r1 Stratagene	5.19e-16
	35	31	1.9	379	224	CEC4671	C.elegans cDNA clone	1.56e-14
	36	31	1.9	379	39	C08467	C.elegans cDNA clone	1.56e-14
	37	28	1.7	280	231	HSA62630	zm02h01.r1 Stratagene	2.89e-10
	38	28	1.7	280	14	AA062630	zm02h01.r1 Stratagene	2.89e-10
	39	26	1.6	331	11	AA054927	SWMFCA800SK Brugia ma	1.40e-07
	40	26	1.6	331	223	BMA54927	SWMFCA800SK Brugia ma	1.40e-07
	41	25	1.5	447	200	G22362	human STS WI-14048.	2.72e-06
	42	25	1.5	447	109	R40271	yf80f04.s1 Homo sapie	2.72e-06
	43	24	1.4	375	185	W18653	mb98b05.r1 Soares mou	4.82e-05

```
   44     24    1.4     375 218  W18653      mb98b05.r1 Soares mou  4.82e-05
   45     24    1.4     450 216  W13185      ma81d02.r1 Soares mou  4.82e-05
```

ALIGNMENTS

```
RESULT    1
LOCUS       T94579        428 bp    mRNA            EST       24-MAR-1995
DEFINITION  ye31g05.s1 Homo sapiens cDNA clone 119384 3'.
ACCESSION   T94579
NID         g728067
KEYWORDS    EST.
SOURCE      human clone=119384 library=Stratagene lung (#937210)
            vector=pBluescript SK- host=SOLR cells (kanamycin resistant)
            primer=-21m13 Rsite1=EcoRI Rsite2=XhoI Normal lung tissue from a 72
            year old male. Cloned unidirectionally. Primer: Oligo dT. Average
            insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:
            5'-GAATTCGGCACGAG-3'; 3' adaptor sequence:
            5'-CTCGAGTTTTTTTTTTTTTTTTT-3'.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 428)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 241
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1..428
                      /organism="Homo sapiens"
                      /clone="119384"
BASE COUNT      112 a    103 c    124 g     76 t     13 others
ORIGIN

  Query Match           20.8%;  Score 345;  DB 151;  Length 428;
  Best Local Similarity  94.3%;  Pred. No. 0.00e+00;
  Matches   383;  Conservative    0;  Mismatches  18;  Indels   5;  Gaps   5;

Db       7 gangggcacaaaccaaagatttattttgcaagtgaaaggggcagcc-aggagacccagaa 65
           || |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Cp    1646 GAGGGGCACAAACCAAAGATTTATTTTGCAAGTGAAAGGGGCAGCCCAGGAGACCCAGAA 1587

Db      66 aaaaggaaggcaaggctgagagcagaaagcctggataaaggaagaccacagaaaggcctg 125
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    1586 AAAAGGAAGGCAAGGCTGAGAGCAGAAAGCCTGGATAAAGGAAGACCACAGAAAGGCCTG 1527

Db     126 caggagccagattgcggccccccagggaaaacccaggaggcaggctgtagagtgatcctgg 185
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    1526 CAGGAGCCAGATTGCGGCCCCCAGGGAAAACCCAGGAGGCAGGCTGTAGAGTGATCCTGG 1467

Db     186 gcccagcctcaaagctggacttggaggggctttan-gactcaattccaggtgcagcattt 244
           |||||||||||||||||   |||||||||||||  |||||||||||||||||||||||||
Cp    1466 GCCCAGCCTCAAAGCTGGGACTGGAGGGGCTTTAGCGACTCAATTCCAGGTGCAGCATTT 1407

Db     245 gcagganttgctgaacaccaggcctntcgggcagcttt-ctgnaacaccngccccgntgc 303
           |||||| |||||||||||||||||| |||||||||||| ||| |||| | |||||| |||
Cp    1406 GCAGGAGTTGCTGAACACCAGGCCTGTCGGGCAGCTTTGCTGGAACAGCCGCCCCGCTGC 1347

Db     304 acaactgtagaagctggaccgttcccgaggnttgggatagagcccatcagcttt-ccctg 362
           ||| |||||||||||||||||||||||| |||||||||||||||||||||||| |||||
Cp    1346 ACAGCTGTAGAAGCTGGACCGTTCCCGAGGATTGGGATAGAGCCCATCAGCTTTGCCCTG 1287

Db     363 gcagaacgtttcttgtncagggcttngggcatgcttcaggttcaga 408
           ||||||||| |||||| |||||||  || |||||| ||||||||||
Cp    1286 GCAGAACGTGTCTTGTCCAGGGCTGGGGCCATGCT-CAGGTTCAGA 1242


RESULT    2
LOCUS       T94272        405 bp    mRNA            EST       24-MAR-1995
DEFINITION  ye31g05.r1 Homo sapiens cDNA clone 119384 5'.
ACCESSION   T94272
NID         g727760
KEYWORDS    EST.
SOURCE      human clone=119384 library=Stratagene lung (#937210)
            vector=pBluescript SK- host=SOLR cells (kanamycin resistant)
            primer=M13RP1 Rsite1=EcoRI Rsite2=XhoI Normal lung tissue from a 72
            year old male. Cloned unidirectionally. Primer: Oligo dT. Average
            insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:
            5'-GAATTCGGCACGAG-3'; 3' adaptor sequence:
            5'-CTCGAGTTTTTTTTTTTTTTTT-3'.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 214
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1..405
                      /organism="Homo sapiens"
                      /clone="119384"
BASE COUNT       64 a    113 c    102 g    115 t     11 others
ORIGIN

  Query Match           14.8%;  Score 245;  DB 151;  Length 405;
  Best Local Similarity  91.6%;  Pred. No. 0.00e+00;
  Matches   371;  Conservative    0;  Mismatches  14;  Indels  20;  Gaps  20;

Db       1 tctgaacctgagcatggccccagccctggacaagacacgttctgccagggcaaagctgat 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1242 TCTGAACCTGAGCATGGCCCCAGCCCTGGACAAGACACGTTCTGCCAGGGCAAAGCTGAT 1301

Db      61 gggctctatcccaatcctcgggaacggtccagcttctacagctgtgcagcggggcg-ctg 119
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Qy    1302 GGGCTCTATCCCAATCCTCGGGAACGGTCCAGCTTCTACAGCTGTGCAGCGGGGCGGCTG 1361

Db     120 ttccagcaaagctgcccgacag-cctggtgttcagcaactcctgcaaatgctgcacctgg 178
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Qy    1362 TTCCAGCAAAGCTGCCCGACAGGCCTGGTGTTCAGCAACTCCTGCAAATGCTGCACCTGG 1421

Db     179 gaattgagtcgctaaagcccctccagtcccagctttgaggctgggcccagggattcactc 238
           |||||||||||||||||||||||||||||||||||||||||||||||||||| || |||||
Qy    1422 -AATTGAGTCGCTAAAGCCCCTCCAGTCCCAGCTTTGAGGCTGGGCCCAGG-AT-CACTC 1478

Db     239 ttacagccttgccttcctggggttttccctgggggccgcaatctgggntccttgcaggc 298
           | ||||||| |||| |||||| |||||||||||| |||||||||| |||| ||||||
Qy    1479 T-ACAGCCT-GCCT-CCTGGG-TTTTCCCTGGGGG-CCGCAATCTGG-CTCCT-GCAGGC 1531
```

```
Db    299 ctttttttgtggtntttcctttantccagggtttnttgttcttcagcnttgcnttcctttt 358
          |||| | ||||| || |||||| |||||| ||| | | ||| |||| |||| ||||||||
Qy   1532 CTTTCT-GTGGTCTT-CCTTTA-TCCAGGCTTTCT-GCTCT-CAGCCTTGCCTTCCTTTT 1586

Db    359 ttnttgggtttcctggggttgcccntttcatttngcaaantaaat 403
          || | |||| ||||||| | |||| ||||| || ||||| |||||
Qy   1587 TTCT-GGGTCTCCTGGGCT-GCCCCTTTCACTT-GCAAAATAAAT 1628


RESULT    3
ID   HSG124     standard; DNA; STS; 241 BP.
AC   G29124;
NI   g1592433
DT   04-OCT-1996 (Rel. 49, Created)
DT   05-OCT-1996 (Rel. 49, Last updated, Version 2)
DE   human STS SHGC-17364.
KW   primer; sequence tagged site; STS sequence.
OS   Homo sapiens (human)
OC   Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC   Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
RN   [1]
RP   1-241
RA   Myers R.M.;
RT   ;
RL   Unpublished.
CC   Contact: Richard M. Myers Stanford Human Genome Center (SHGC)
CC   Stanford University School of Medicine Department of Genetics,
CC   M-344, Stanford, CA 94305, USA Tel: 4157259687 Fax: 4157259689
CC   Email: myers@shgc.stanford.edu Primer A: AAAGCCTGGATAAAGGAAGACC
CC   Primer B: GCTGCACCTGGAATTGAGTC STS size: 150 PCR Profile: Initial
CC   incubation: 94 degrees C for 90 seconds Denaturation: 94 degrees C
CC   for 15 seconds Annealing: 62 degrees C for 23 seconds
CC   Polymerization: 72 degrees C for 30 seconds PCR Cycles: 30 Thermal
CC   Cycler: Perkin Elmer 9600 Protocol: Template: 25 ng Primer: each 1
CC   uM dNTPs: each 200 uM Taq Polymerase: 0.05 units/ul Total Vol: 10
CC   ul Buffer: MgCl2: 2.5 mM KCl: 50 mM Tris-HCl: 20 mM pH: 8.3
CC   Prepared with primer pairs derived from T94579--Merck/UniEST.
FH   Key             Location/Qualifiers
FH
FT   source          1..241
FT                   /organism="Homo sapiens"
FT                   /map="1"
FT   STS             91..240
FT   primer_bind     91..112
FT   primer_bind     complement(221..240)
SQ   Sequence 241 BP; 78 A; 53 C; 77 G; 30 T; 3 other;

  Query Match           13.0%;  Score 215;  DB 250;  Length 241;
  Best Local Similarity  97.0%;  Pred. No. 0.00e+00;
  Matches   230;  Conservative    0;  Mismatches   5;  Indels   2;  Gaps   2;

Db      7 gangggcacaaaccaaagatttattttgcaagtgaaaggggcagcc-aggagacccagaa 65
          || ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Cp   1646 GAGGGGCACAAACCAAAGATTTATTTTGCAAGTGAAAGGGGCAGCCCAGGAGACCCAGAA 1587

Db     66 aaaaggaaggcaaggctgagagcagaaagcctggataaaggaagaccacagaaaggcctg 125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1586 AAAAGGAAGGCAAGGCTGAGAGCAGAAAGCCTGGATAAAGGAAGACCACAGAAAGGCCTG 1527

Db    126 caggagccagattgcggcccccagggaaaacccaggaggcaggctgtagagtgatcctgg 185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1526 CAGGAGCCAGATTGCGGCCCCCAGGGAAAACCCAGGAGGCAGGCTGTAGAGTGATCCTGG 1467

Db    186 gcccagcctcaaagctggacttggaggggctttan-gactcaattccaggtgcagca 241
          |||||||||||||||||   |||||||||||||  ||||||||||||||||||||
Cp   1466 GCCCAGCCTCAAAGCTGGGACTGGAGGGGCTTTAGCGACTCAATTCCAGGTGCAGCA 1410


RESULT    4
LOCUS       G29124        241 bp    DNA             STS       04-OCT-1996
DEFINITION  human STS SHGC-17364.
ACCESSION   G29124
NID         g1592433
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 241)
  AUTHORS   Myers,R.M.
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157259687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu

            Primer A: AAAGCCTGGATAAAGGAAGACC
            Primer B: GCTGCACCTGGAATTGAGTC
            STS size: 150
            PCR Profile:
                    Initial incubation: 94 degrees C for 90 seconds

                    Denaturation:       94 degrees C for 15 seconds
                    Annealing:          62 degrees C for 23 seconds
                    Polymerization:     72 degrees C for 30 seconds
                    PCR Cycles:         30
                    Thermal Cycler:     Perkin Elmer 9600
            Protocol:
                    Template:       25 ng
                    Primer:         each 1 uM
                    dNTPs:          each 200 uM
                    Taq Polymerase: 0.05 units/ul
                    Total Vol:      10 ul

            Buffer:
                    MgCl2:          2.5 mM
                    KCl:            50  mM
                    Tris-HCl:       20  mM
                    pH:             8.3

            Prepared with primer pairs derived from T94579--Merck/UniEST.
FEATURES             Location/Qualifiers
     source          1..241
                     /organism="Homo sapiens"
                     /map="1"
     STS             91..240
     primer_bind     91..112
     primer_bind     complement(221..240)
BASE COUNT       78 a     53 c     77 g     30 t      3 others
ORIGIN

  Query Match           13.0%;  Score 215;  DB 202;  Length 241;
  Best Local Similarity  97.0%;  Pred. No. 0.00e+00;
  Matches   230;  Conservative    0;  Mismatches   5;  Indels   2;  Gaps   2;

Db      7 gangggcacaaaccaaagatttattttgcaagtgaaaggggcagcc-aggagacccagaa 65
          || ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Cp   1646 GAGGGGCACAAACCAAAGATTTATTTTGCAAGTGAAAGGGGCAGCCCAGGAGACCCAGAA 1587

Db     66 aaaaggaaggcaaggctgagagcagaaagcctggataaaggaagaccacagaaaggcctg 125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1586 AAAAGGAAGGCAAGGCTGAGAGCAGAAAGCCTGGATAAAGGAAGACCACAGAAAGGCCTG 1527

Db    126 caggagccagattgcggcccccagggaaaacccaggaggcaggctgtagagtgatcctgg 185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1526 CAGGAGCCAGATTGCGGCCCCCAGGGAAAACCCAGGAGGCAGGCTGTAGAGTGATCCTGG 1467

Db    186 gcccagcctcaaagctggacttggaggggctttan-gactcaattccaggtgcagca 241
          |||||||||||||||||   |||||||||||||  ||||||||||||||||||||
Cp   1466 GCCCAGCCTCAAAGCTGGGACTGGAGGGGCTTTAGCGACTCAATTCCAGGTGCAGCA 1410
```

```
RESULT    5
LOCUS       H43270        405 bp    mRNA            EST       31-JUL-1995
DEFINITION  yp05a12.r1 Homo sapiens cDNA clone 186526 5' similar to gb:M80927
            CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   H43270
NID         g919322
KEYWORDS    EST.
SOURCE      human clone=186526 library=Soares breast 3NbHBst vector=pT7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=M13RP1 Rsite1=Not I Rsite2=Eco RI Adult human.
            1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
            digested with Not I and cloned into the Not I and Eco RI sites of a
            modified pT7T3 vector (Pharmacia). Library went through one round
            of normalization to a Cot = 20. Library constructed by Bento Soares
            and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 323
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1..405
                     /organism="Homo sapiens"
                     /clone="186526"
BASE COUNT      106 a    112 c     95 g     86 t      6 others
ORIGIN

  Query Match             4.8%;  Score 79;  DB 41;  Length 405;
  Best Local Similarity  63.0%;  Pred. No. 1.44e-100;
  Matches   196;  Conservative    0;  Mismatches 114;  Indels   1;  Gaps   1;

Db      1 gctgcttcccagatgcccttgaccgcttccnctgtacccacatcatctacagctttgcca 60
          ||| | | ||  | | | | |||| |  ||  || |||||| ||||||||  ||| ||
Qy    145 GCTTCCTGCCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTG 204

Db     61 atataagcaacgatcacatcgacacctgggagtggaatgatgtgacgctctacggcatgc 120
            || | |||| | ||  |   ||||   |||||||||| | ||| ||||||     |
Qy    205 GCATGACCAACCACCAGCTGAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGT 264

Db    121 tcaacacactcaagaacaggaaccccaacctgaagactctcttgtctgtgggaggatgga 180
          ||||    || ||||| | ||| ||||| |||||||| || ||  |  | ||||| ||||
Qy    265 TCAATGGCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGA 324

Db    181 ncnttgggtctcaaagattttccaagatagcctccaacacccagagtcgccggactttca 240
            |  |  |||| |  ||  |  | || |   |||   || | |  || | ||| ||
Qy    325 ATTTCAGCACTCAGAAGTTCACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTG 384

Db    241 tcaagtcagtaccgccatttctgcgcacccatggctttgatggggctggaccttgcctgg 300
          |||| || |        ||||||||||   |  |||||| ||  || ||||||| ||||
Qy    385 TCAACTCGGCCATCAGGTTTCTGCGCAAATACAGCTTTGACGGC-CTTGACCTTGACTGG 443

Db    301 gtctaccctgg 311
          |  ||||| ||
Qy    444 GAGTACCCAGG 454


RESULT    6
LOCUS       N75653        541 bp    mRNA            EST       02-APR-1996
DEFINITION  yv29c01.r1 Homo sapiens cDNA clone 244128 5' similar to gb:M80927
            CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   N75653
NID         g1238231
KEYWORDS    EST.
SOURCE      human clone=244128 primer=reverse ET library=Soares fetal liver
            spleen 1NFLS vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Pac I Rsite2=Eco RI Liver
            and spleen from a 20 week-post conception male fetus. 1st strand
            cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 441
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1..541
                     /organism="Homo sapiens"
                     /clone="244128"
     mRNA            <1..>541
BASE COUNT      149 a    134 c    117 g    136 t      5 others
ORIGIN

  Query Match             4.7%;  Score 78;  DB 113;  Length 541;
  Best Local Similarity  66.7%;  Pred. No. 1.34e-98;
  Matches   166;  Conservative    0;  Mismatches  82;  Indels   1;  Gaps   1;

Db     27 ctgggcaggtgtagtggtcttgctgcttctccagggaggatctgcctacaaactggtttg 86
          ||||||||||  |  ||||| |||||  |  |||   | || |||||    |||||||| ||
Qy     44 CTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTG 103

Db     87 ctactttaccaactggtcccaggaccggcagaaccaggaaaa-ttcacccctgagaatat 145
          |||||| |||||||| ||||| || | |||     |||     |||   ||  || |  |
Qy    104 CTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTT 163

Db    146 tgacccctttcctatgctctcatctcatctattcattcgccagcatcgaaaacaacaaggt 205
          ||||||  ||| ||| | || ||||||||  | |||||  ||||    ||| || || |
Qy    164 GGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCT 223

Db    206 tatcatcaaggacaagagtgaagtgatgctctaccagaccatcaacagtctcaaaaccaa 265
```

```
          | || ||  ||   || ||| | ||  |||||||||    ||||  | || || |  |
Qy    224 GAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGAT 283

Db    266 gaatcccaa 274
          |||||||||
Qy    284 GAATCCCAA 292


RESULT    7
LOCUS       N75653         541 bp    mRNA            EST       02-APR-1996
DEFINITION  yv29c01.r1 Homo sapiens cDNA clone 244128 5' similar to gb:M80927
            CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   N75653
NID         g1238231
KEYWORDS    EST.
SOURCE      human clone=244128 primer=reverse ET library=Soares fetal liver
            spleen 1NFLS vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Pac I Rsite2=Eco RI Liver
            and spleen from a 20 week-post conception male fetus. 1st strand
            cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 441
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.

            NCBI gi: 1238231
FEATURES              Location/Qualifiers
     source           1..541
                      /organism="Homo sapiens"
                      /clone="244128"
                      /note="human"
     mRNA             <1..>541
BASE COUNT      149 a    134 c    117 g    136 t      5 others
ORIGIN

  Query Match             4.7%;  Score 78;  DB 207;  Length 541;
  Best Local Similarity  66.7%;  Pred. No. 1.34e-98;
  Matches   166;  Conservative    0;  Mismatches  82;  Indels   1;  Gaps   1;

Db     27 ctgggcaggtgtagtggtcttgctgcttctccagggaggatctgcctacaaactggtttg 86
          |||||||||| |  ||||| ||||| |  |||   | || |||||    |||||||| ||
Qy     44 CTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTG 103

Db     87 ctactttaccaactggtcccaggaccggcagaaccaggaaaa-ttcacccctgagaatat 145
          |||||| |||||||| ||||| || | |||    |||     |||   ||  || |  |
Qy    104 CTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTT 163

Db    146 tgaccccttcctatgctctcatctcatctattcattcgccagcatcgaaaacaacaaggt 205
          ||||||  ||| ||| | || ||||||||  | |||||  ||||    ||| || || |
Qy    164 GGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCT 223

Db    206 tatcatcaaggacaagagtgaagtgatgctctaccagaccatcaacagtctcaaaaccaa 265
          | || ||  ||   || ||| | ||  |||||||||    ||||  | || || |  |
Qy    224 GAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAAGAT 283

Db    266 gaatcccaa 274
          |||||||||
Qy    284 GAATCCCAA 292


RESULT    8
LOCUS       T65854         533 bp    mRNA            EST       20-FEB-1995
DEFINITION  yc24b03.s1 Homo sapiens cDNA clone 81581 3' similar to SP:S27879
            S27879 SECRETORY PROTEIN YM-1 PRECURSOR - ;.
ACCESSION   T65854
NID         g674899
KEYWORDS    EST.
SOURCE      human clone=81581 library=Stratagene lung (#937210)
            vector=pBluescript SK- host=SOLR cells (kanamycin resistant)
            primer=-21m13 Rsite1=EcoRI Rsite2=XhoI Normal lung tissue from a 72
            year old male. Cloned unidirectionally. Primer: Oligo dT. Average
            insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:
            5'-GAATTCGGCACGAG-3'; 3' adaptor sequence:
            5'-CTCGAGTTTTTTTTTTTTTTTT-3'.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High qality sequence stops: 307
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1..533
                      /organism="Homo sapiens"
                      /clone="81581"
BASE COUNT      140 a    123 c    136 g    128 t      6 others
ORIGIN

  Query Match             4.4%;  Score 73;  DB 144;  Length 533;
  Best Local Similarity  70.1%;  Pred. No. 7.80e-89;
  Matches   131;  Conservative    0;  Mismatches  55;  Indels   1;  Gaps   1;

Db    250 tggacaagagggtaaggnccctggttgcaggatttnccagtgaagtcatccatgtcaata 309
          ||||  || |||||  | ||||||||||||||    || |  |||||||| | ||| |
Cp   1162 TGGATGAGGGGGTATCGGCCCTGGTTGCAGGAGAAGCCGGCAAAGTCATCTAAGTCCAGT 1103

Db    310 gaccagatcatggctcctcccaggtttaaattctttaagaactgaaccttggtctccata 369
          | ||||| |||||| || |||||        | ||| |    |     |||||||| |  |
Cp   1102 GCCCAGACCATGGCCCCGCCCAGTCCCTTCTGCTTCAGATAGCTGACCTTGGTTTTGAAG 1043

Db    370 ctcttcacatcatcatagcccacccacttggttccccttgactncgtaggggacctgctg 429
          |||| |||||||||| |||||||||||| |||| ||  ||    |||||| |||| |||
Cp   1042 CTCTCCACATCATCAAAGCCCACCCACT-GGTTGTCCCGGAAGATGTAGGGCACCTTCTG 984

Db    430 atcctgg 436
```

```
          |||||||
Cp    983 ATCCTGG 977


RESULT    9
LOCUS       G24226         533 bp    DNA             STS       31-MAY-1996
DEFINITION  human STS WI-14165.
ACCESSION   G24226
NID         g1344552
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human STSs derived from sequences in dbEST and the Unigene
            collection.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: AGGGCAACAAGTCAACAGCT
            Primer B: TCTCCCAGAATTCTCATGTGG
            STS size: 132
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Derived from dbEST (genbank accession T65854).
FEATURES              Location/Qualifiers
     source           1..533
                      /organism="Homo sapiens"
                      /map="461.5 cR from top of Chr1 linkage group"
     STS              34..165
     primer_bind      34..53
     primer_bind      complement(145..165)
BASE COUNT      140 a    123 c    136 g    128 t      6 others
ORIGIN

  Query Match             4.4%;  Score 73;  DB 200;  Length 533;
  Best Local Similarity  70.1%;  Pred. No. 7.80e-89;
  Matches   131;  Conservative    0;  Mismatches  55;  Indels   1;  Gaps   1;

Db    250 tggacaagagggtaaggncctggttgcaggatttnccagtgaagtcatccatgtcaata 309
          ||||  || |||||  | ||||||||||||||    || |  |||||||| | ||| |
Cp   1162 TGGATGAGGGGGTATCGGCCCTGGTTGCAGGAGAAGCCGGCAAAGTCATCTAAGTCCAGT 1103

Db    310 gaccagatcatggctcctcccaggtttaaattctttaagaactgaaccttggtctccata 369
          | ||||| |||||| || |||||       | ||| |   |    |||||||| |  |

Cp   1102 GCCCAGACCATGGCCCCGCCCAGTCCCTTCTGCTTCAGATAGCTGACCTTGGTTTTGAAG 1043

Db    370 ctcttcacatcatcatagcccacccacttggttccccttgactncgtaggggacctgctg 429
          |||| |||||||||| |||||||||||| ||||  ||  ||    |||||| |||| |||
Cp   1042 CTCTCCACATCATCAAAGCCCACCCACT-GGTTGTCCCGGAAGATGTAGGGCACCTTCTG 984

Db    430 atcctgg 436
          |||||||
Cp    983 ATCCTGG 977


RESULT   10
LOCUS       N40107         311 bp    mRNA            EST       22-JAN-1996
DEFINITION  yw72b09.r1 Homo sapiens cDNA clone 257753 5' similar to gb:M80927
            CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   N40107
NID         g1163652
KEYWORDS    EST.
SOURCE      human clone=257753 primer=T7 library=Soares placenta 8to9weeks
            2NbHP8to9W vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI two
            placentae: one from 8 weeks and another from 9 weeks post
            conception. 1st strand cDNA was primed with a Not I - oligo(dT)
            primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M. Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 268
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1..311
                      /organism="Homo sapiens"
                      /clone="257753"
     mRNA             <1..>311
BASE COUNT       93 a     81 c     68 g     68 t      1 others
ORIGIN

  Query Match             4.3%;  Score 71;  DB 79;  Length 311;
  Best Local Similarity  65.8%;  Pred. No. 5.83e-85;
  Matches   156;  Conservative    0;  Mismatches  80;  Indels   1;  Gaps   1;

Db     69 ctgggcaggtgtagtggtcttgctgcttctccagggaggatctgcctacaaactggtttg 128
          |||||||||| |  ||||| ||||| |  |||   | || |||||    |||||||| ||
Qy     44 CTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTG 103

Db    129 ctactttaccaactggtcccaggaccggcagaaccaggaaaa-ttcacccctgagaatat 187
          |||||| |||||||| ||||| || | |||    |||     |||   ||  || | |
Qy    104 CTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTT 163
```

```
Db    188 tgacccctcctatgctctcatctcatctattcattcgccagcatcgaaaacaacaaggt 247
          ||||||  ||| ||| | || |||||||  | |||||  ||||    ||| || || |
Qy    164 GGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCT 223

Db    248 tatcatcaaggacaagagtgaagtgatgctctaccagaccatcaacagtctcnaaaa 304
          | || ||  ||   || ||| | ||  |||||||||    ||||  | ||  | ||
Qy    224 GAGCACCACTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATGGCCTGAAGAA 280


RESULT   11
LOCUS       W51302        336 bp    mRNA            EST       29-MAY-1996
DEFINITION  ma16f08.r1 Life Tech mouse brain Mus musculus cDNA clone 304743 5'
            similar to gb:X93035 M.musculus mRNA for BRP39 protein (MOUSE);.
ACCESSION   W51302
NID         g1339742
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 336)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:201503
            Seq primer: ETPrimer
            High quality sequence stop: 148.
FEATURES              Location/Qualifiers
     source           1..336
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Vector: pCMV-SPORT2; Site_1: SalI; Site_2: NotI;
                      Cloned unidirectionally.  Primer: Oligo dT. pCMV-SPORT2
                      vector."
                      /clone="304743"
                      /clone_lib="Life Tech mouse brain"
                      /dev_stage="adult"
                      /lab_host="DH10B"
     mRNA             <1..>336
BASE COUNT       83 a    100 c     89 g     64 t
ORIGIN

  Query Match             4.3%;  Score 71;  DB 169;  Length 336;
  Best Local Similarity  69.2%;  Pred. No. 5.83e-85;
  Matches   137;  Conservative    0;  Mismatches  60;  Indels   1;  Gaps   1;

Db    140 caggctttgcggtcctgatgctgctccagagctgctctgcgtacaagttggtctgctact 199
          ||||  ||   ||||||| || || |||   |  |||||||    ||  ||||||||||||
Qy     49 CAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTGCTACT 108

Db    200 acaccaactggtcccagtaccgggaaggcaatgggagctgcttcccagacgccctcgacc 259
          ||||||||||  |||||||| |  | ||  | |   ||| | | ||  | | | | ||||
Qy    109 TCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACC 168

Db    260 attccctgtgcacccatatcatctacagctttgcaa-catcagccacaacaagctcagca 318
              ||| ||||||||  ||||||||  ||| ||   ||| | | || || |||| ||||
Qy.   169 CCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAACCACCAGCTGAGCA 228

Db    319 catccgagtggaatgacg 336
          |  | ||||||||||||||
Qy    229 CCACTGAGTGGAATGACG 246


RESULT   12
LOCUS       H24821        266 bp    mRNA            EST       07-JUL-1995
DEFINITION  yl42f01.r1 Homo sapiens cDNA clone 160921 5' similar to gb:M80927
            CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   H24821
NID         g893720
KEYWORDS    EST.
SOURCE      human clone=160921 library=Soares breast 3NbHBst vector=pT7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=M13RP1 Rsite1=Not I Rsite2=Eco RI Adult human.
            1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
            digested with Not I and cloned into the Not I and Eco RI sites of a
            modified pT7T3 vector (Pharmacia). Library went through one round
            of normalization to a Cot = 20. Library constructed by Bento Soares
            and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 266)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 84
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1..266
                      /organism="Homo sapiens"
                      /clone="160921"
BASE COUNT       58 a     80 c     70 g     51 t      7 others
ORIGIN

  Query Match             4.0%;  Score 66;  DB 35;  Length 266;
  Best Local Similarity  68.3%;  Pred. No. 2.27e-75;
  Matches   127;  Conservative    0;  Mismatches  58;  Indels   1;  Gaps   1;

Db     14 ggaccacagtccatagaatcctcggccagcaggtcccctatgccaccaagggcaaccagt 73
          || ||||    || |||||||  |  ||| |||| |||||   |  |  || ||||||||
Qy    958 GGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCCTACATCTTCCGGGACAACCAGT 1017

Db     74 gggtaggatacgacgaccaggaaagcgtcaaaagcaaggtgcagtacctgaaggacaggc 133
          |||| || |  || ||   ||| ||| |||||| ||||||    || |||||| | | |
Qy   1018 GGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAGGTCAGCTATCTGAAGCAGAAGG 1077

Db    134 agctgg-cgggcgcatggtatgggccctngacctggatnacttncagggctccttctgcg 192
            |||| |||| |||||| ||||| || ||| | ||| ||||    |||| || ||||
Qy   1078 GACTGGGCGGGGCCATGGTCTGGGCACTGGACTTAGATGACTTTGCCGGCTTCTCCTGCA 1137

Db    193 gccagg 198
            ||||||
Qy   1138 ACCAGG 1143
```

```
RESULT   13
LOCUS       R55530        396 bp    mRNA            EST       22-MAY-1995
DEFINITION  yj79c12.r1 Homo sapiens cDNA clone 154966 5' similar to gb:M80927
            CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   R55530
NID         g824825
KEYWORDS    EST.
SOURCE      human clone=154966 library=Soares breast 2NbHBst vector=pT7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=M13RP1 Rsite1=Not I Rsite2=Eco RI Adult female.
            1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
            digested with Not I and cloned into the Not I and Eco RI sites of a
            modified pT7T3 vector (Pharmacia). Library went through one round
            of normalization to a Cot = 230. Library constructed by Bento
            Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 396)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 284
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1..396
                     /organism="Homo sapiens"
                     /clone="154966"
BASE COUNT       88 a    106 c    109 g     87 t      6 others
ORIGIN

  Query Match             3.9%;  Score 65;  DB 134;  Length 396;
  Best Local Similarity  68.3%;  Pred. No. 1.81e-73;
  Matches   114;  Conservative    0;  Mismatches  53;  Indels   0;  Gaps   0;

Db    125 caggctttgtggtcctggtgctgctccagtgctgctctgcatacaaactggtctgctact 184
          |||| ||  |||||||| || || |||  ||  |||||||    ||||||||||||||||
Qy     49 CAGGTTTCATGGTCCTGCTGATGATCCCATGGGGCTCTGCTGCAAAACTGGTCTGCTACT 108

Db    185 acaccagctggtcccagtaccgggnangcgatgggagctgcttcccagatgcccttgacc 244
          ||||| |||| |||||||| |     | || |   ||| | | ||  | | | | ||||
Qy    109 TCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCGCTTCCTGCCCAAGGACTTGGACC 168

Db    245 gnntcctctgtacccacatcatctacagctttgccaatataagcaac 291
              ||| || |||||| ||||||||  ||| ||    || | ||||
Qy    169 CCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCATGACCAAC 215


RESULT   14
LOCUS       W10705        306 bp    mRNA            EST       26-APR-1996
DEFINITION  ma47g07.r1 Soares mouse p3NMF19.5 Mus musculus cDNA 5' similar to
            gb:M80927 CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   W10705
NID         g1285179
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 306)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA+ET
            High quality sequence stop: 172.

            NCBI gi: 1285179
FEATURES             Location/Qualifiers
     source          1..306
                     /organism="Mus musculus"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo."
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1..>306
BASE COUNT       78 a     76 c     85 g     67 t
ORIGIN

  Query Match             3.6%;  Score 60;  DB 215;  Length 306;
  Best Local Similarity  72.1%;  Pred. No. 4.65e-64;
  Matches    98;  Conservative    0;  Mismatches  38;  Indels   0;  Gaps   0;

Db     78 agaaggttcccttcgctaccaagggcaaccagtgggtggggtatgaggacaaggagagtg 137
          ||||||| |||| |     |  || ||||||||||||||| | ||| ||   ||||||
Qy    985 AGAAGGTGCCCTACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCT 1044

Db    138 tcaaaaacaaggttgggttcctgaaggagaagaagctggcaggagcatcggtgtgggcac 197
          |||||| ||||||  | |  |||||| |||||   ||||  || ||   ||| |||||||
Qy   1045 TCAAAACCAAGGTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCAC 1104

Db    198 tggatttggatgattt 213
          |||| || ||||| ||
Qy   1105 TGGACTTAGATGACTT 1120


RESULT   15
LOCUS       W10705        306 bp    mRNA            EST       05-SEP-1996
DEFINITION  ma47g07.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 313884 5'
            similar to gb:M80927 CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
ACCESSION   W10705
NID         g1285179
KEYWORDS    EST.
```

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 306)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA+ET
            High quality sequence stop: 172.
FEATURES              Location/Qualifiers
     source           1..306
                      /organism="Mus musculus"
                      /note="Vector: pT7T3D (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by Bento
                      Soares and M.Fatima Bonaldo. RNA was kindly provided by
                      Dr. Minoru Ko (Wayne State University)."
                      /clone="313884"
                      /clone_lib="Soares mouse p3NMF19.5"
                      /dev_stage="19.5 dpc total fetus"
                      /lab_host="DH10B (ampicillin resistant)"
     mRNA             <1..>306
BASE COUNT       78 a     76 c     85 g     67 t
ORIGIN

  Query Match             3.6%;  Score 60;  DB 174;  Length 306;
  Best Local Similarity  72.1%;  Pred. No. 4.65e-64;
  Matches    98;  Conservative    0;  Mismatches  38;  Indels   0;  Gaps   0;

Db     78 agaaggttcccttcgctaccaagggcaaccagtgggtggggtatgaggacaaggagagtg 137
          ||||||| |||| |     |  || ||||||||||||||| | ||| ||   ||||||
Qy    985 AGAAGGTGCCCTACATCTTCCGGGACAACCAGTGGGTGGGCTTTGATGATGTGGAGAGCT 1044

Db    138 tcaaaaacaaggttgggttcctgaaggagaagaagctggcaggagcatcggtgtgggcac 197
          |||||| |||||| | |  |||||| |||||   ||||  || ||   ||| |||||||
Qy   1045 TCAAAACCAAGGTCAGCTATCTGAAGCAGAAGGGACTGGGCGGGGCCATGGTCTGGGCAC 1104

Db    198 tggatttggatgattt 213
          |||| || ||||| ||
Qy   1105 TGGACTTAGATGACTT 1120


Search completed: Sun Jun 29 21:08:58 1997
Job time : 1195 secs.
```

****

****

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 27 16:57:44 1997; MasPar time 11.53 Seconds
440.291 Million cell updates/sec
Tabular output not generated.

Title: >US-08-663-618A-4
Description: (1-466) from US08663618A.pep
Perfect Score: 3440
Sequence: 1 MVRSVAWAGFMVLLMIPWGS.....QQSCPTGLVFSNSCKCCTWN 466

Scoring table: PAM 150
Gap 11

Searched: 92623 seqs, 10896596 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq26
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19

Statistics: Mean 36.091; Variance 158.898; scale 0.227

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1344	39.1	668	14	R73993	Hamster oviduct speci	2.56e-114
2	1339	38.9	537	14	R73991	Bovine oviduct specif	7.77e-114
3	1326	38.5	718	14	R73992	Murine oviduct specif	1.39e-112
4	906	26.3	321	1	P81342	Polypeptide involved	2.43e-72
5	597	17.4	371	13	R70025	Tobacco chitinase enc	3.72e-43
6	600	17.4	377	13	R70029	Tobacco cluster-A pro	1.95e-43
7	584	17.0	424	12	R63441	Trichoderma harzianum	6.10e-42
8	579	16.8	389	10	R56861	Aphanocladium album m	1.79e-41
9	579	16.8	423	6	R33072	Aphanocladium album p	1.79e-41
10	572	16.6	389	6	R33068	Aphanocladium album m	8.03e-41
11	466	13.5	866	19	W02159	Soluble chitinase.	5.32e-31
12	392	11.4	561	12	R64823	Chitinase A.	3.09e-24
13	375	10.9	561	9	R48981	Bacterial chitinase-A	1.07e-22
14	271	7.9	238	6	R32547	HaNPV IE-1 gene prote	1.95e-13
15	220	6.4	1046	19	W02156	Periplasmic chitodext	4.80e-09
16	139	4.0	25	13	R70751	YKL-40 N-terminal pep	1.90e-02
17	139	4.0	25	12	R70745	YKL-40 N-terminal seq	1.90e-02
18	136	4.0	29	11	R60073	Bovine Oviduct Glycop	3.22e-02
19	100	2.9	398	18	R95856	S. pyogenes speB gene	1.39e+01
20	99	2.9	428	6	R31372	Murine bone morphogen	1.63e+01
21	99	2.9	428	16	R86905	Murine BMP-9.	1.63e+01
22	99	2.9	553	14	R80445	Murine soluble epoxid	1.63e+01
23	101	2.9	958	2	R07575	Glucoamylase encoded	1.18e+01
24	101	2.9	958	1	P81181	Sequence of glucoamyl	1.18e+01
25	101	2.9	1070	4	R21521	Alpha galactosidase.	1.18e+01
26	95	2.8	438	9	R47259	Pre-pro-VGR1.	3.05e+01
27	92	2.7	369	18	R98125	Human lymphocyte cell	4.85e+01
28	92	2.7	1356	5	R26999	Novel type III RTK en	4.85e+01
29	91	2.6	19	12	R70746	YKL-40 internal pepti	5.66e+01
30	91	2.6	19	13	R70752	YKL-40 internal pepti	5.66e+01
31	90	2.6	256	17	R97235	MuTu putative oncogen	6.59e+01
32	90	2.6	459	17	R88058	Protein encdoed by Mu	6.59e+01
33	88	2.6	513	9	R47264	Pre-pro BMP6.	8.92e+01
34	88	2.6	513	2	R07311	Human Bone Morphogene	8.92e+01
35	88	2.6	513	6	R32904	Human BMP-6 peptide s	8.92e+01
36	88	2.6	513	7	R36735	Human BMP-6.	8.92e+01
37	90	2.6	575	3	P50018	Sequence encoded by t	6.59e+01
38	90	2.6	709	3	P50017	Sequence encoded by t	6.59e+01
39	90	2.6	709	2	P70423	FixD nitrogen fixatio	6.59e+01
40	88	2.6	896	17	R92750	Human EGF receptor su	8.92e+01
41	87	2.5	281	18	R94503	1-acylglycerol-3-phos	1.04e+02
42	87	2.5	281	16	R87724	Full length meadowfoa	1.04e+02
43	87	2.5	284	19	R99248	Limnanthes 2-acyltran	1.04e+02
44	87	2.5	542	16	R90772	Bacillus stearothermo	1.04e+02
45	87	2.5	2351	3	P50319	Human antihaemophilic	1.04e+02

ALIGNMENTS

RESULT 1
ID R73993 standard; Protein; 668 AA.
AC R73993;
DT 10-JAN-1996 (first entry)
DE Hamster oviduct specific glycoprotein.
KW Hamster oviduct specific glycoprotein; recombinant production;
KW HOGP.
OS Cricetulus griseus.
FH Key Location/Qualifiers
FT Peptide 1..18
FT /label= sig_peptide
FT Peptide 19..668
FT /label= mat_peptide
PN J07107979-A.
PD 25-APR-1995.
PF 15-AUG-1994; 214227.
PR 19-AUG-1993; JP-227881.
PA (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR WPI; 95-190179/25.
DR N-PSDB; Q90444.
PT New DNA encoding an oviduct-specific glycoprotein - useful for
PT recombinant protein production in high quantities.
PS Claim 4; Pages 16-18; 22pp; Japanese.
CC Q90444 encodes R73993 hamster oviduct specific glycoprotein (HOGP).
CC The cDNA can be used for the commercial recombinant prodn. of
CC HOGP in high quantities.
SQ Sequence 668 AA;

Query Match 39.1%; Score 1344; DB 14; Length 668;
Best Local Similarity 48.3%; Pred. No. 2.56e-114;
Matches 185; Conservative 84; Mismatches 103; Indels 11; Gaps 9;

```
Db      4 wvg-lvllmkpndgtayklvcyftnwahsrpvpasilprdldpflcthlifafasmsnnq 62
          | | :|||| |  |:| ||||||||||: |   | :||:|||| ||||||:|||:|:|:|
Qy      7 WAGFMVLLMIPW-GSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQ 65

Db     63 ivannlqdekilypefnklkernralktllsvggwnfgtsrfttmlstlasrekfigsvv 122
          : : :  ||  || ||| ||  |  |||||::|||||:| :|| |::|  :|: |: | :
Qy     66 LSTTEWNDET-LYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAI 124

Db    123 sflrthgfdgldlfflypglrgspindrwnflflieelqfafekealltqrprlllsaav 182
            |||  :|||||   ||| :|||  |:  |  |:::|  ||: ||  : : ||||||||
Qy    125 RFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAV 184
```

```
Db    183 -sgipyiiqtsydvhllgrrldfinvlsydlhgswekstghnsplfslpedpkssa-fam 240
          :|  |: :::|:|  ::: |||:|:::|):|||||| |||||||:   |:: ::| : :
Qy    185 PAGQTYV-DAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNV 243

Db    241 n--y--wrnlgapadkllmgfpaygrtfhllreskng1qaasmgpaspgkytkqagflay 296
          :    |   |:|| ||::| |:|||:| |   |   : |:: |:::|| :||::| |||
Qy    244 DAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAY 303

Db    297 yevcsfiqraekhwidhqyvpyaykgkewvgyddavsfsykamfvkkehfggamvwtldm 356
          |||||    | |::|: | ||| :: ::|||:|| || |  ::|   :||||||:||:
Qy    304 YEVCSW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDL 362

Db    357 ddvrgtfcgngpfplvhilnell 379
          || |  |  | :||:: | : |
Qy    363 DDFAGFSCNQGRYPLIQTLRQEL 385


RESULT    2
ID   R73991 standard; Protein; 537 AA.
AC   R73991;
DT   10-JAN-1996  (first entry)
DE   Bovine oviduct specific glycoprotein.
KW   Bovine oviduct specific glycoprotein; recombinant production;
KW   BOGP.
OS   Bos taurus.
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT   /label= sig_peptide
FT   Peptide         19..537
FT   /label= mat_peptide
PN   J07107979-A.
PD   25-APR-1995.
PF   15-AUG-1994; 214227.
PR   19-AUG-1993; JP-227881.
PA   (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR   WPI; 95-190179/25.
DR   N-PSDB; Q90442.
PT   New DNA encoding an oviduct-specific glycoprotein - useful for
PT   recombinant protein production in high quantities.
PS   Claim 4; Pages 7-9; 22pp; Japanese.
CC   Q90442 encodes R73991 bovine oviduct specific glycoprotein (BOGP).
CC   The cDNA can be used for the commercial recombinant prodn. of
CC   BOGP in high quantities.
SQ   Sequence   537 AA;

  Query Match           38.9%;  Score 1339;  DB 14;  Length 537;
  Best Local Similarity  50.1%;  Pred. No. 7.77e-114;
  Matches   190;  Conservative   81;  Mismatches 101;  Indels   7;  Gaps   6;

Db      6 glllvlkhhdgaahklvcyftnwafsrpgpasilprdldpflcthlvfafasmsnnqivp 65
          |::::|:   |:| |||||||||  | | | :|:|:|||| |||||::|||:|:|:|:
Qy      9 GFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLST 68

Db     66 kdpqdekilypefnklkernrglktllsiggwnfgtvrfttmlstfsnrerfvssviall 125
          :  ||  || ||| ||  |  |||||:|||||:| :|| |::| :||: ||:| | :|
Qy     69 TEWNDET-LYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFL 127

Db    126 rthgfdgldlfflypglrgspardrwtfvflleellqafkneaqltmrprlllsaavsgd 185
          |  :||||||   ||| :|||| |:  |  |:::|  ||  ||| : : |||||||::
Qy    128 RKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG 187

Db    186 phvvqkayearllgrlldfisvlsydlhgswekvtghnsplfs-l--pgdpks-sa-yam 240
          |: :||   ::: |||:::::||:|||||||||||||:    :| : | :   |:
Qy    188 QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAV 247

Db    241 nywrqlgvppekllmglptygrtfhllkasqnelraqavgpaspgkytkqagflayyeic 300
          | | | |: ||::|:|||||:| |  :|:  : | | |:::|| :||::| |||||:|
Qy    248 QQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVC 307

Db    301 cfvrrakkrwindqyvpyafkgkewvgyddaisfgykaffikrehfggamvwtldlddfr 360
          : | |::| || ||| |: ::|||:|| ||  |  ::|:  :|||||:||||||
Qy    308 SW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFA 366
```

```
Db    361 gyfcgtgpfplvhtlnnll 379
          |: |  | :||::||   |
Qy    367 GFSCNQGRYPLIQTLRQEL 385


RESULT    3
ID   R73992 standard; Protein; 718 AA.
AC   R73992;
DT   10-JAN-1996  (first entry)
DE   Murine oviduct specific glycoprotein.
KW   Murine oviduct specific glycoprotein; recombinant production;
KW   MOGP.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT   /label= sig_peptide
FT   Peptide         19..718
FT   /label= mat_peptide
PN   J07107979-A.
PD   25-APR-1995.
PF   15-AUG-1994; 214227.
PR   19-AUG-1993; JP-227881.
PA   (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR   WPI; 95-190179/25.
DR   N-PSDB; Q90443.
PT   New DNA encoding an oviduct-specific glycoprotein - useful for
PT   recombinant protein production in high quantities.
PS   Claim 4; Pages 11-14; 22pp; Japanese.
CC   Q90443 encodes R73992 murine oviduct specific glycoprotein (MOGP).
CC   The cDNA can be used for the commercial recombinant prodn. of
CC   MOGP in high quantities.
SQ   Sequence   718 AA;

  Query Match           38.5%;  Score 1326;  DB 14;  Length 718;
  Best Local Similarity  49.0%;  Pred. No. 1.39e-112;
  Matches   190;  Conservative   82;  Mismatches 107;  Indels   9;  Gaps   6;

Db      7 lvllmkhsdgtayklvcyftnwahsrpgpasimphdldpflcthlifafasmsnnqivak 66
          :||||    |:| ||||||||||: | | | ::| |||| ||||||:|||:|:|:|: :
Qy     11 MVLLMIPW-GSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTT 69

Db     67 nlqdenvlypefnklkernrelktllsiggwnfgtsrftamlstlanrekfidsvisflr 126
          :  ||  || ||| ||  |  |||||:|||||:| :|| |::|  ||: |::| | |||
Qy     70 EWNDET-LYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLR 128

Db    127 ihgfdgldlfflypglrgspphdrwnflflieelqfaferealltqhprlllsaavsgip 186
          :||||||   ||| :|||: |:  |  |:::|  ||::||  :   |||||||::
Qy    129 KYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQ 188

Db    187 siihtsydalllgrrldfinvlsydlhgswekftghnsplfslpedskssa--y---amn 241
          : : ::|:   ::: |||:|:::||:|||||| |||||||:   |:| ::|      |:
Qy    189 TYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQ 248

Db    242 ywrklgtpadklimgfptygrnfyllkesknglqtasmgpaspgkytkqagflayyevcs 301
          |   |||| |||:| |||||:| |   |   : ::: |:::|| :||::| ||||||||
Qy    249 QWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCS 308

Db    302 fvqrakkhwidyqyvpyafkgkewlgyddtisfsykamyvkrehfggamvwtldmddvrg 361
              | |::|: | ||| |: ::|:|:|:||  ||  |  |:|:  :|||||:||:||  |
Qy    309 W-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAG 367

Db    362 tfcgngpfplvhil-nellvqtesnstp 388
          |  | :||:: |  || :   :::||
Qy    368 FSCNQGRYPLIQTLRQELSLPYLPSGTP 395


RESULT    4
ID   P81342 standard; protein; 321 AA.
AC   P81342;
DT   19-OCT-1990  (first entry)
DE   Polypeptide involved in protective mechanisms
KW   Immune response; cell growth.
PN   J63032898-A.
```

```
PD   30-JAN-1988.
PF   16-JUL-1986; 167518.
PR   16-JUL-1986; JP-167518.
PA   (DAIN) Dainippon Pharm KK.
DR   WPI; 88-068419/10.
DR   P-PSDB; P81342.
PT   New polypeptide and DNA encoding it - related to protective
PT   mechanisms such as immune response etc.
PS   Disclosure; p; Japanese.
CC   This polypeptide is involved in protective mechanisms such as immune
CC   response, cell growth and activation of protective functions.
SQ   Sequence   321 AA;

  Query Match           26.3%;  Score 906;  DB 1;  Length 321;
  Best Local Similarity  49.1%;  Pred. No. 2.43e-72;
  Matches   113;  Conservative   66;  Mismatches  45;  Indels   6;  Gaps   3;

Db       7 tgfvvlvllqccsayklvcyytswsqyregdgscfpdaldrflcthiiysfanisndhid 66
           :||:||:::   || |||||:|:|:|||:|::  :|  ||  ||||:||:|| ::| ::
Qy       8 AGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLS 67

Db      67 twewndvtlygmlntlknrnpnlktllsvggwnfgsqrfskiasntqsrrtfiksvppfl 126
           | |||| |||  :| ||: ||:|||||::||||||:|:|: : : : :|:||::|   ||
Qy      68 TTEWNDETLYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFL 127

Db     127 rthgfdgvdlawlypg-rg----dkqhfttlikemkaefikeaq-pgkkqlllsaalsag 180
           |  :|||:|| | ||| :|    ||::|||: ::   |  ||| :|| :|||||::||
Qy     128 RKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG 187

Db     181 kvtidssydiakisqhldlisimtydfhgawrgttghhsplfrgqedasp 230
             :|::|:: ||:|:||::::|:|||||:|   |||:|||:: ||::::
Qy     188 QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGA 237


RESULT    5
ID   R70025 standard; Protein; 371 AA.
AC   R70025;
DT   29-SEP-1995  (first entry)
DE   Tobacco chitinase encoded by tobacco Cluster-A cDNA clone cA-3.
KW   Tobacco; chitinase; antifungal; fungicide; Cluster-A.
OS   Nicotiana tabacum.
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT   /label= signal peptide
FT   Misc_difference 22
FT   /note= "not present in genomic clone"
FT   Misc_difference 182..195
FT   /note= "Determined by protein sequencing"
FT   Misc_difference 216..241
FT   /note= "Determined by protein sequencing"
PN   EP-639642-A.
PD   22-FEB-1995.
PF   17-AUG-1993; 202425.
PR   17-AUG-1993; EP-202425.
PA   (MOGE-) MOGEN INT NV.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PI   Apotheker-de Groot M,  Bol JF,  Cornelissen BJC,  Linthorst HJM;
PI   Melchers LS,  Ponstein AS,  Sela-buurlage MB;
DR   WPI; 95-083454/12.
DR   N-PSDB; Q82973.
PT   New plant protein having endo-chitinase activity - used in
PT   antifungal compsns. and to develop transformed plants which are
PT   less susceptible to fungal infection.
PS   Claim 3; Page 19-20; 43pp; English.
CC   Screening of a lambda ZAP cDNA library of TMV-infected Samsunn NN
CC   tobacco plants with a probe derived from PROB40 (a partial Cluster-A
CC   cDNA clone) resulted in the isolated of 11 positively hybridising
CC   clones. Analysis revealed that all were identical and corresp. to
CC   Cluster-A cDNA. The nt sequence of clone cA-3 is given in Q82973/
CC   R70025. This clone lacks 7 codons of the 5' part of the ORF. The
CC   cDNA clone was completed by performing a PCR reaction on clone cA-3
CC   with primer T7 (Q82794) and primer P1 (Q82975). A partial AA
CC   sequence of the protein isolated from TMV-infected tobacco leaves
CC   was obtd. and found to be almost identical to the AA sequence
CC   deduced from the cDNA.
SQ   Sequence   371 AA;

  Query Match           17.4%;  Score 597;  DB 13;  Length 371;
  Best Local Similarity  33.0%;  Pred. No. 3.72e-43;
  Matches   107;  Conservative   86;  Mismatches 108;  Indels  23;  Gaps  18;

Db      36 nnidstlfthlfcafadlnpqsnqliispenqds-fsqftstvqrknpsvktflsiaggr 94
           :::|::| |||: ||| :   ::||  :  |::: : :| : : ::|| :||:|:| ||:
Qy      44 KDLDPSLCTHLIYAFAGMT--NHQLSTTEWNDETLYQEFNG-LKKMNPKLKTLLAI-GGW 99

Db      95 a-dttaygimarqpnsrksfidssirlarqfgfhgldldweyp--l-s-atdmtnlgill 149
             |  :  |   :|:| :|::|:||: | ::| ||||||||    | | |:  :  |:
Qy     100 NFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLV 159

Db     150 newrtainmearnsgraallltaavsysprvnglnypvesvarnlnwinlmaydfygpnw 209
             :  |:  ||: ||:  |||:|||: :       | |: :|:||: :|||||| |: |
Qy     160 QDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGS-W 218

Db     210 spsqtnshaqlfdpvnh--isgs---dg-inawiqagvptkklvlgipfygyawrlvnpn 263
             |  :: |:   :    ::|   |: :  |:| | |: ||:||:| || :  | :::
Qy     219 EKV-TGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSS 277

Db     264 ihdlrapaagksnvga-vddgsm-tynrirdyivqtrattvynativgdycysgsnwisy 321
           : |||:| :  |:   :|:| :|  :      |:   : :   |  | :  : |:::
Qy     278 DTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQR-IQDQK-VP-YIFRDNQWVGF 334

Db     322 ddtqsvrnkvnyvkgrgllgyfaw 345
           || :| : ||:|:| :|| |   |
Qy     335 DDVESFKTKVSYLKQKGLGGAMVW 358


RESULT    6
ID   R70029 standard; Protein; 377 AA.
AC   R70029;
DT   29-SEP-1995  (first entry)
DE   Tobacco cluster-A protein encoded by genomic clone.
KW   Tobacco; chitinase; antifungal; fungicide; Cluster-A.
OS   Nicotiana tabacum.
FH   Key             Location/Qualifiers
FT   Peptide         1..25
FT   /label= signal peptide
FT   Misc_difference 28..29
FT   /note= "cDNA encodes additional Ile here"
FT   Misc_difference 188..201
FT   /label= Determined by sequencing the protein
FT   Misc_difference 222..247
FT   /note= "Determined by sequencing the protein"
PN   EP-639642-A.
PD   22-FEB-1995.
PF   17-AUG-1993; 202425.
PR   17-AUG-1993; EP-202425.
PA   (MOGE-) MOGEN INT NV.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PI   Apotheker-de Groot M,  Bol JF,  Cornelissen BJC,  Linthorst HJM;
PI   Melchers LS,  Ponstein AS,  Sela-buurlage MB;
DR   WPI; 95-083454/12.
DR   N-PSDB; Q82976.
PT   New plant protein having endo-chitinase activity - used in
PT   antifungal compsns. and to develop transformed plants which are
PT   less susceptible to fungal infection.
PS   Claim 3; Page 25-28; 43pp; English.
CC   Screening of a lambda ZAP cDNA library of TMV-infected Samsunn NN
CC   tobacco plants with a probe derived from PROB40 (a partial Cluster-A
CC   cDNA clone) resulted in the isolated of 11 positively hybridising
CC   clones. Analysis revealed that all were identical and corresp. to
CC   Cluster-A cDNA.The sequence of cDNA clone cA-3 is given in Q82973/
CC   R70025. A genomic library of N. tabacum was screened using the
CC   Cluster-A cDNA insert of clone cA-3 as a probe (see Q82977, Q82978).
CC   The complete nt. sequence of Cluster-A cDNA including the deduced
CC   primary structure of the Cluster-A protein the 5' and 3' UTR regions
CC   of the gene are shown in Q82976/R70029. Comaprison of the cDNA
```

```
CC   clone with the Cluster-A gene revealed that these sequences share
CC   a high degree of identity (94%). The Cluster-A precursor protien
CC   contains a putative signal peptide as well as 4 potential N-linked
CC   glycosylation sites (N-X-S/T). The predicted mature protein has
CC   a calculated mol. wt. of 39,033 Da.
SQ   Sequence   377 AA;

  Query Match           17.4%;  Score 600;  DB 13;  Length 377;
  Best Local Similarity  33.3%;  Pred. No. 1.95e-43;
  Matches   108;  Conservative   87;  Mismatches 106;  Indels  23;  Gaps  18;

Db     42 nnidstlfthlfcafadlnpqlnqliispenqds-frqftstvqrknpsvktflsiaggr 100
          :::|::| |||: ||| :  :  || :  |::: :::| : : ::|| :||:|:| ||:
Qy     44 KDLDPSLCTHLIYAFAGMTNH--QLSTTEWNDETLYQEFNG-LKKMNPKLKTLLAI-GGW 99

Db    101 ansta-ygimarqpnsrksfidssirlarqlgfhgldldweyp--l-saa-dmtnlgtll 155
            ||  :  |   :|:| :|::|:||: |  :| |||||||||    |:| |:  : ||:
Qy    100 NFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLV 159

Db    156 newrtaintearnsgraallltaavsnsprvnglnypveslarnldwinlmaydfygpnw 215
           :   |:  ||: ||:  |||:|||: :       | |: :|:||| :|||||| |: |
Qy    160 QDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGS-W 218

Db    216 spsqtnshaqlfdpvnh--vsgs---dg-inawiqagvptkklvlgipfygyawrlvnan 269
             |  :: |:  :    ::|   |: :  |:| | |: ||:||:| || :  | :::
Qy    219 EKV-TGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSS 277

Db    270 ihglrapaagksnvga-vddgsm-tynrirdyivesrattvynativgdycysgsnwisy 327
             : |||:| :  |:   :|:| :|  :      ::   : :   |  | :  : |:::
Qy    278 DTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQR-IQDQK-VP-YIFRDNQWVGF 334

Db    328 ddtqtvrnkvnyvkgrgllgyfaw 351
          || :: : ||:|:| :|| |   |
Qy    335 DDVESFKTKVSYLKQKGLGGAMVW 358


RESULT    7
ID   R63441 standard; Protein; 424 AA.
AC   R63441;
DT   06-JUL-1995  (first entry)
DE   Trichoderma harzianum P1 endochitinase.
KW   Trichoderma harzianum P1; endochitinase; inhibition;
KW   chitin containing fungi; herbivorous insects.
OS   Trichoderma harzianum.
PN   WO9424288-A.
PD   27-OCT-1994.
PF   10-FEB-1994; U01198.
PR   14-APR-1993; US-045269.
PA   (CORR ) CORNELL RES FOUND INC.
PI   Harman GE,  Hayes CK,  Lorito M,  Tronsmo A;
DR   WPI; 94-341867/42.
DR   N-PSDB; Q78144.
PT   Trichoderma harzianum P1 endo:chitinase and chitobiase genes -
PT   useful in biological control of chitin-contg. fungi and insects
PT   (Eng).
PS   Example 8; Page 45; 62pp; English.
CC   Q78144 encodes R63441 the Trichoderma harzianum P1 endochitinase,
CC   which can be used to inhibit chitin containing fungi and chitin
CC   containing herbivorous insects, especially from the genera below
CC   Fusarium, Gliocladium, Rhizoctonia, Trichoderma, Uncinula, Ustilage,
CC   Erysiphe, Botrytis, Saccharomyces, Sclerotium and Alternaria.
SQ   Sequence   424 AA;

  Query Match           17.0%;  Score 584;  DB 12;  Length 424;
  Best Local Similarity  36.4%;  Pred. No. 6.10e-42;
  Matches   103;  Conservative   68;  Mismatches  95;  Indels  17;  Gaps  13;

Db    112 qlfk-lkkangnlkvmlsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidv 169
          | |: ||| | :|| :|:|||| || :|    :|  ||  | ::|| |::  :|||:|:
Qy     78 QEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 137

Db    170 dweyp-addtqatnmvlllkeirsqldayaaqy-apg-yhfllsiaapagpehy-sflhm 225
          ||||| :: : | ::  :   ::   :|:  :  ::|  ::||| | |||     :   :

Qy    138 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 197

Db    226 sdlgqvldyvnlmaydyagswssysghdanlfanpsnpnssp-yntdqaikdyikggvpa 284
            ::| ||:|||||||: |||   :|||:: |:    :: :::  | | |: : :  | ||
Qy    198 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 257

Db    285 skivlgmpiygraf--estgg--igqtysgigs-gsw--engiwdy-kvlpkagatvqyd 336
          ||::|||| |||:|   |::   :|   :| |: |:   | |:  |  | :  ||| |
Qy    258 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQ-- 315

Db    337 svaqayysydpsskelisfdtpdmintkvsylknlglggsmfw 379
             : :  :|    :: ::||  : ::|||||||  ||||:| |
Qy    316 RIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT    8
ID   R56861 standard; Protein; 389 AA.
AC   R56861;
DT   17-FEB-1995  (first entry)
DE   Aphanocladium album mature endochitinase.
KW   oxalic acid degradation; oxalate oxidase; selection marker;
KW   filamentous fungus endochitinase; precursor;
KW   pathogen resistance; plant cell selection.
OS   Aphanocladium album.
FH   Key             Location/Qualifiers
FT   mat_peptide     1..1167
FT   /*tag= a
FT   /product= endochitinase
PN   WO9413790-A.
PD   23-JUN-1994.
PF   07-DEC-1993; F01203.
PR   07-DEC-1992; FR-014721.
PA   (SNFI ) ELF SANOFI.
PA   (ERAP ) SOC NAT ELF AQUITAINE.
PI   Grezes-Besset B,  Grison R,  Pignard A,  Schneider M;
DR   WPI; 94-217870/26.
DR   N-PSDB; Q68353.
PT   Use of gene encoding oxalic acid degrading enzyme - for plant
PT   cell selection, esp. gene coupled to gene-encoding protein
PT   conferring pathogen resistance
PS   Disclosure and Claim 8; Page 39-40; 64pp; French.
CC   The sequence coding for wheat germin, a protein induced during
CC   germination of wheat, is a preferred example of DNA coding for a
CC   protein able to degrade oxalic acid. Such DNA can be coupled to a
CC   sequence of interest, e.g. one coding for enzymes which confer
CC   resistance to pathogens (esp. endochitinases). Plant calli which
CC   have been successfully transformed by the sequences are selected
CC   by their ability to grow on oxalic acid-contg. medium. R56861 is
CC   a known sequence of a mature endochitinase; the sequence was
CC   previously disclosed in EP-531218.
SQ   Sequence   389 AA;

  Query Match           16.8%;  Score 579;  DB 10;  Length 389;
  Best Local Similarity  36.6%;  Pred. No. 1.79e-41;
  Matches   102;  Conservative   64;  Mismatches  95;  Indels  18;  Gaps  16;

Db     81 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 139
          ||| | ::| :|:|||| || :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy     83 LKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 142

Db    140 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 194
           :::: | ::  :   |:  ::: :  : |  ::||| | ||| : |    :  : ::
Qy    143 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 201

Db    195 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 253
          ||::|||||: ||| : |||:: :|   ::: |::  | | ||| :  | ||:|::
Qy    202 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 261

Db    254 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 305
          |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    262 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 321

Db    306 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 344
```

```
          | :    :: ::||    : |||| || |||||:| |
Qy    322 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT    9
ID   R33072 standard; Protein; 423 AA.
AC   R33072;
DT   30-JUN-1993  (first entry)
DE   Aphanocladium album pre-pro-endochitinase.
KW   filamentous fungus; pathogen resistance.
OS   Aphanocladium album.
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT   /label= signal
FT   /note= "pre-sequence"
FT   Peptide         23..34
FT   /note= "pro-sequence"
FT   Protein         35..423
FT   /label= endochitinase
PN   EP-531218-A.
PD   10-MAR-1993.
PF   04-SEP-1992; 402414.
PR   06-SEP-1991; FR-011072.
PA   (SNFI ) ELF SANOFI.
PA   (ERAP ) SOC NAT ELF AQUITAINE.
PI   Blaiseau P,  Legoux R,  Leguay J,  Schneider M;
DR   WPI; 93-078700/10.
DR   N-PSDB; Q37711, Q37712.
PT   Recombinant DNA encoding fungal endo:chitinase - useful for
PT   protecting plants against fungi, insects, etc.
PS   Claim 1; Fig 1; 75pp; French.
CC   Clone CH3C was isolated from a cDNA bank prepared from A.album
CC   mRNA by screening with anti-chitinase antiserum. Three probes were
CC   designed based on the CH3C sequence, corresponding to the start,
CC   middle and downstream regions of the coding sequence. The probes
CC   were used to isolate the full-length coding sequence of A.album
CC   endochitinase from an A.album cDNA library in pTZ19R. A 1.6kb
CC   fragment was identified in several positive clones. The sequence
CC   of the fragment contained an ORF encoding a 423 amino acid protein.
CC   The mature endochitinase can be used to control pathogens
CC   such as fungi, bacteria, arthropods and nematodes on plants.
SQ   Sequence   423 AA;

  Query Match           16.8%;  Score 579;  DB 6;  Length 423;
  Best Local Similarity  36.6%;  Pred. No. 1.79e-41;
  Matches   102;  Conservative   64;  Mismatches  95;  Indels  18;  Gaps  16;

Db    115 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 173
          ||| | ::| :|:|||| || :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy     83 LKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 142

Db    174 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 228
           :::: | ::  :   |:   ::: :  : |  ::||| | ||| : |    :  : ::
Qy    143 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 201

Db    229 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 287
            ||::||||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    202 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 261

Db    288 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 339
          |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    262 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 321

Db    340 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 378
          | :    :: ::||    : |||| || |||||:| |
Qy    322 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT   10
ID   R33068 standard; Protein; 389 AA.
AC   R33068;
DT   30-JUN-1993  (first entry)
DE   Aphanocladium album mature endochitinase.
KW   filamentous fungus; pathogen resistance.
OS   Aphanocladium album.
PN   EP-531218-A.
PD   10-MAR-1993.
PF   04-SEP-1992; 402414.
PR   06-SEP-1991; FR-011072.
PA   (SNFI ) ELF SANOFI.
PA   (ERAP ) SOC NAT ELF AQUITAINE.
PI   Blaiseau P,  Legoux R,  Leguay J,  Schneider M;
DR   WPI; 93-078700/10.
DR   N-PSDB; Q37704.
PT   Recombinant DNA encoding fungal endo:chitinase - useful for
PT   protecting plants against fungi, insects, etc.
PS   Claim 22; Page 60-61; 75pp; French.
CC   This mature protein from the filamentous fungus Aphanocladium
CC   album has endochitinase activity and is used to control pathogens
CC   such as fungi, bacteria, arthropods and nematodes on plants.
CC   Recombinant DNA comprising the endochitinase coding sequence is
CC   specifically intended for transforming Nicotiana tabacum, Helianthus
CC   annuus and Brassica napus to pathogen resistance.
SQ   Sequence   389 AA;

  Query Match           16.6%;  Score 572;  DB 6;  Length 389;
  Best Local Similarity  36.2%;  Pred. No. 8.03e-41;
  Matches   101;  Conservative   65;  Mismatches  95;  Indels  18;  Gaps  16;

Db     81 lkkqnrnmkvmvsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 139
          ||| | ::| :::|||| || :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy     83 LKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 142

Db    140 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 194
           :::: | ::  :   |:   ::: :  : |  ::||| | ||| : |    :  : ::
Qy    143 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 201

Db    195 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 253
            ||::||||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    202 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 261

Db    254 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 305
          |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    262 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 321

Db    306 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 344
          | :    :: ::||    : |||| || |||||:| |
Qy    322 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT   11
ID   W02159 standard; Protein; 866 AA.
AC   W02159;
DT   14-JAN-1997  (first entry)
DE   Soluble chitinase.
KW   Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;
KW   Beta-N-acetylglucosamidase; chitin; oligosaccharide; catabolic;
KW   catabolism.
OS   Vibrio furnissii.
PN   WO9625424-A1.
PD   22-AUG-1996.
PF   13-FEB-1996; U02332.
PR   13-FEB-1995; US-386727.
PA   (UYJO ) UNIV JOHNS HOPKINS.
PI   Bassler B,  Chitlaru E,  Keyhani N,  Roseman S,  Rowe C;
PI   Yu C;
DR   WPI; 96-393335/39.
DR   N-PSDB; T36390.
PT   Chitin biosynthetic enzymes end I, exo I and exo II  - are
PT   periplasmic chito:dextrinase(s), periplasmic beta-GlcNAcidase(s) and
PT   aryl beta-N-acetyl:gluco:amidase(s), respectively
PS   Example 4; Page 79-82; 101pp; English.
CC   Periplasmic chitodextrinase (W02156), periplasmic
CC   Beta-N-acetylglucosaminidase (W02157) and aryl
CC   Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin
CC   oligosaccharides with the structure (GlcNAc)n where n is 2 or
```

```
CC   higher, by contacting them with soluble chitin.  The enzymes are
CC   encoded by the genes endI, exoI and exoII respectively.  They are
CC   all genes involved in the catabolic pathway of chitin.
SQ   Sequence   866 AA;

  Query Match           13.5%;  Score 466;  DB 19;  Length 866;
  Best Local Similarity  31.0%;  Pred. No. 5.32e-31;
  Matches    94;  Conservative   80;  Mismatches 104;  Indels  25;  Gaps  20;

Db    253 fgnlmalkkanpnlkilpsvggwtlsdp-fy-ffsdktkrdtfvasmkeylqtwkffdgv 310
          : :: :||| ||:|| | ::||| :|   |    :   :|:||| |   :|:   | ||:
Qy     77 YQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSF-DGL 135

Db    311 didwefpggqganpnlggpndgatyvalmkelramldeleaet-grqyeltsaisaggdk 369
          |:|||:||:||: | :   ::   : :|: :|   ::: ||:| |::  | ||  ::|:
Qy    136 DLDWEYPGSQGS-PAVD--KE--RFTTLVQDLANAFQQ-EAQTSGKERLLLSAAVPAGQT 189

Db    370 iakvdyqaaq--qymdyiflmsydfsgafdlknlahqtnly--asswdpatkyttdkgvk 425
              |:    | :|:: ||:||| |: : |  :| : ||       :|:    | :|
Qy    190 YVDAGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNSPLYKRQEESGAAASLNVDAAVQ 248

Db    426 allgqgvtpgkvvvgaamygrgwt-gvng-yqagnpftgsat-gpikgtwengvvdyrdi 482
            |  |  ::|:::| : |||: | : ::  : | | |||:| ||:    | |:: | ::
Qy    249 QWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFT-K-EGGMLAYYEV 306

Db    483 vnnrmgagweqgydetaeapyvfkastgdlisfdndrsvkakgqyvlanqlgglfaweid 542
           : ::||   |  ::    ||:|: :  : ::||:  | |:|  |:  : |||   | :|
Qy    307 CS-WKGAT-KQRIQD-QKVPYIFRDN--QWVGFDDVESFKTKVSYLKQKGLGGAMVWALD 361

Db    543 adn 545
            |:
Qy    362 LDD 364


RESULT   12
ID   R64823 standard; Protein; 561 AA.
AC   R64823;
DT   23-JUL-1995  (first entry)
DE   Chitinase A.
KW   Chitinase A; protein secretion; biological control agent;
KW   transgenic plant; chitin; plant pathogen; nematode; fungus;
KW   crop protection; pCHIT1251.
OS   Serratia marcescens QMB1466.
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT   /label= Sig_peptide
PN   US5374540-A.
PD   20-DEC-1994.
PF   26-MAR-1984; 593691.
PR   26-MAR-1984; US-593691.
PR   18-JUL-1986; US-888033.
PR   09-JUL-1990; US-550253.
PA   (DNAP ) DNA PLANT TECHNOLOGY CORP.
PI   Jones JDG,  Suslow T;
DR   WPI; 95-035648/05.
DR   N-PSDB; Q76290.
PT   Plant cell contg. signal sequence of chitinase A gene - for
PT   secretion of foreign polypeptide, esp. chitinase A for protection
PT   against pathogenic fungi and nematodes.
PS   Disclosure; Fig. 1A-1C; 20pp; English.
CC   Serratia marcescens QMB1466 genomic DNA was partially digested,
CC   ligated into vector pLAFR1, packaged into lambda phage particles,
CC   and used to transfect Escherichia coli DH1 (ATCC 33849) cells.
CC   Plating on chitin medium was used to screen for chitinase activity.
CC   Clone C12 (ATCC 67152) was isolated.  Its sequence is given in
CC   Q76290, and the deduced protein sequence in R64823.
SQ   Sequence   561 AA;

  Query Match           11.4%;  Score 392;  DB 12;  Length 561;
  Best Local Similarity  28.5%;  Pred. No. 3.09e-24;
  Matches    92;  Conservative   80;  Mismatches 126;  Indels  25;  Gaps  18;

Db    231 pfaalqkaqkgvtawddpykgnfgqlmalkqahpdlkilpsiggwtlsdpfff-mgdkvk 289
          :||:: : | : | |:|     : :: :||  :| || | :|||| :|   | |    :
Qy     57 AFAGMTNHQLSTTEWNDE-T-LYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATAN 114

Db    290 -rdrfvgsvkeflqtwkffdgvdidwefpggkganpnlgspqdgetyvllmkelramldq 348
           |: || |   ||:   | ||:|:|||:||: |: | :    | | :  |: :|   ::|
Qy    115 NRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV----DKERFTTLVQDLANAFQQ 168

Db    349 lsaetgr-kyeltsaisagkdkidkvaynvaq--nsmdhiflmsydfygafdlknlghqt 405
           :  :|: :  |::|::||   :|  :|:|     ::| : ||:||| |: : |  || :
Qy    169 EAQTSGKERLLLSAAVPAGQTYVD-AGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNS 226

Db    406 alnar---pgsrhrlhhgerrecaagqgvkpgkivvgtamygrgwtgvngyqnnipftgt 462
          :|  |   :|:   |:     :     |  ::|:::| : |||: |  :: :  :   :|
Qy    227 PLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPAT 286

Db    463 hravkgtw--engivdyrqiasqfmsgewqytydataeapyvfkpstgdlitsddarsvq 520
            :  |    | |:: | :: |  :::  |   :     ||:|:    : :  ||  |
Qy    287 GSGTPGPFTKEGGMLAYYEVCS-WKGATKQ-RIQD-QKVPYIFR-DN-QWVGFDDVESFK 341

Db    521 akgkyvldkqlgglfsweidadn 543
          :|  |: :| |||    | :| |:
Qy    342 TKVSYLKQKGLGGAMVWALDLDD 364


RESULT   13
ID   R48981 standard; Protein; 561 AA.
AC   R48981;
DT   04-SEP-1994  (first entry)
DE   Bacterial chitinase-A fragment from pCHIT1251.
KW   Chitinase-A; enzyme; transgenic plant; biological control;
KW   disease-resistance; crop improvement; Agrobacterium; fungus;
KW   nematode; insect.
PN   US5290687-A.
PD   01-MAR-1994.
PF   26-MAR-1984; 593691.
PR   26-MAR-1984; US-593691.
PR   18-JUL-1986; US-888033.
PR   09-JUL-1990; US-550253.
PR   14-AUG-1992; US-930970.
PA   (DNAP ) DNA PLANT TECHNOLOGY CORP.
PI   Jones JDG,  Suslow T;
DR   WPI; 94-074348/09.
DR   N-PSDB; Q57966.
PT   Inhibition of chitinous fungal plant pathogens - by transforming
PT   plant with DNA encoding chitinase activity
PS   Disclosure; Fig.1; 13pp; English.
CC   Plasmid pCHIT1251 containing the chitinase-A gene (chiA), fused
CC   directly to a nos promoter,  may be used to transform, via
CC   Agrobacterium sp., a crop plant.  The resulting transgenic plants
CC   expressing the chitinase-A protein fragment are protected against
CC   fungi, nematodes, insects and disease agents. The protein is
CC   preferably expressed by the plant only at the site of pathogen
CC   attack, e.g. only in the root cells.  The plants express the
CC   chitinase in a biologically active form.
SQ   Sequence   561 AA;

  Query Match           10.9%;  Score 375;  DB 9;  Length 561;
  Best Local Similarity  27.9%;  Pred. No. 1.07e-22;
  Matches    90;  Conservative   83;  Mismatches 125;  Indels  25;  Gaps  18;

Db    231 pfaalqkaqkgvtawddpykgnfgqlmalkqahpdlkilpsiggwtlsdpfff-mgdkvk 289
          :||:: : | : | |:|     : :: :||  :| || | :|||| :|   | |    :
Qy     57 AFAGMTNHQLSTTEWNDE-T-LYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATAN 114

Db    290 -rdrfvgsvkeflqswkffdgvdidwefpggkganpnlgspqdgetyvlvmkevgamldq 348
           |: || |   ||:   | ||:|:|||:||: |: | :    | | :  :: :::  ::|
Qy    115 NRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV----DKERFTTLVQDLANAFQQ 168

Db    349 lwaetgr-kyeltsaisagkdkidkvaynvaq--nsmdhiflmsydfygafdlknlghqt 405
              :|: :  |::|::||   :|  :|:|     ::| : ||:||| |: : |  || :
Qy    169 EAQTSGKERLLLSAAVPAGQTYVD-AGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNS 226

Db    406 alnar---pgsrhrlhhgegrecaagqgvkpgkivvgtamygrgwtgvngyqnnipftgt 462
```

```
          :|  |   :|:   |:   : :     |  ::|:::| : |||: |  :: :  :   :|
Qy    227 PLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPAT 286

Db    463 hravkgtw--engivdyrqiasqfmsgewqytydgtaeapyvfkpstgdlitsddarsvq 520
            :  |    | |:: | :: |  :::  |   :     ||:|:    : :  ||  |
Qy    287 GSGTPGPFTKEGGMLAYYEVCS-WKGATKQ-RIQD-QKVPYIFR-DN-QWVGFDDVESFK 341

Db    521 akgkyvldkqlgglfsweidadn 543
          :|  |: :| |||   | :| |:
Qy    342 TKVSYLKQKGLGGAMVWALDLDD 364


RESULT   14
ID   R32547 standard; Protein; 238 AA.
AC   R32547;
DT   17-JUN-1993  (first entry)
DE   HaNPV IE-1 gene protein (partial sequence).
KW   Heliothis armigera nuclear polyhedrosis virus; Ha; NPV;
KW   insecticide; immediate early-1.
OS   Heliothis armigera nuclear polyhedrosis virus.
PN   WO9303144-A.
PD   18-FEB-1993.
PF   05-AUG-1992; AU0413.
PR   05-AUG-1991; AU-007576.
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PI   Christian PD, Dall DJ, Gordon KHJ, Hanzlik TN, Sriskantha A.
DR   WPI; 93-076501/09.
DR   N-PSDB; Q35982.
PT   Insect virus with reduced capacity to occlude viral particles - used
PT   for controlling proliferation of insect pests without horizontal
PT   transmission
PS   Disclosure; Fig 2; 51pp; English.
CC   The sequence is that encoded by the Heliothis armigera (Ha)
CC   nuclear polyhedrosis virus (NPV) immediate early-1 (IE-1) gene
CC   (partial sequence).
SQ   Sequence   238 AA;

  Query Match            7.9%;  Score 271;  DB 6;  Length 238;
  Best Local Similarity  34.1%;  Pred. No. 1.95e-13;
  Matches    58;  Conservative   43;  Mismatches  51;  Indels  18;  Gaps  14;

Db     44 erdkflsnpgwrefvltwkf-fdgididwefpggkganpnvgdverdnntyiallgelra 102
          :|: |:::  | | | :|: |||:|:|||:||: |: | | | ||    : :|: :|
Qy    115 NRQTFVNSAI-R-F-L-RKYSFDGLDLDWEYPGSQGS-PAV-DKER----FTTLVQDLAN 164

Db    103 mldq-vqiqtnrtlelttaisagidkiaav-ncdraqqyldkifvmsydfkgawsntdlg 160
           ::|  |   :  | |::|::||   : |  : |:  | || : :|:||| |:| :   |
Qy    165 AFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVT-G 223

Db    161 hqtaly--gsawkpnepytanvavdallaqrvnp-kklvlgvamygrgwt 207
          | ::||       :  :   : ||:  | |:  |  ||:||:: |||: |
Qy    224 HNSPLYKRQEESGAAASLNVDAAVQQWL-QKGTPASKLILGMPTYGRSFT 272


RESULT   15
ID   W02156 standard; Protein; 1046 AA.
AC   W02156;
DT   14-JAN-1997  (first entry)
DE   Periplasmic chitodextrinase.
KW   Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;
KW   Beta-N-acetylglucosamidase; chitin; oligosaccharide; catabolic;
KW   catabolism.
OS   Vibrio furnissii.
PN   WO9625424-A1.
PD   22-AUG-1996.
PF   13-FEB-1996; U02332.
PR   13-FEB-1995; US-386727.
PA   (UYJO ) UNIV JOHNS HOPKINS.
PI   Bassler B,  Chitlaru E,  Keyhani N,  Roseman S,  Rowe C;
PI   Yu C;
DR   WPI; 96-393335/39.
DR   N-PSDB; T36387.
PT   Chitin biosynthetic enzymes end I, exo I and exo II  - are
PT   periplasmic chito:dextrinase(s), periplasmic beta-GlcNAcidase(s) and
PT   aryl beta-N-acetyl:gluco:amidase(s), respectively
PS   Claim 2; Page 68-71; 101pp; English.
CC   Periplasmic chitodextrinase (W02156), periplasmic
CC   Beta-N-acetylglucosaminidase (W02157) and aryl
CC   Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin
CC   oligosaccharides with the structure (GlcNAc)n where n is 2 or
CC   higher, by contacting them with soluble chitin.  The enzymes are
CC   encoded by the genes endI, exoI and exoII respectively.  They are
CC   all genes involved in the catabolic pathway of chitin.
SQ   Sequence   1046 AA;

  Query Match            6.4%;  Score 220;  DB 19;  Length 1046;
  Best Local Similarity  31.1%;  Pred. No. 4.80e-09;
  Matches    52;  Conservative   44;  Mismatches  56;  Indels  15;  Gaps  11;

Db    432 vadggf-ytmttnadgsinqqgietfadsavemmrkyrfdgldidlrisnidggtgnpdd 490
          :| ||  ::    :|        :|| :||:  :||| |||||:| :  :  |: |:|
Qy     94 LAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD-W--EYPGSQGSPA- 149

Db    491 tafsesrraylmnsyhelmrvlrekldvasaqdgvhymltiaapss-ayllrgmetmavt 549
              | | : |:   ::|   :::  : :|: :  : :|: | |:: :|:  | |   ::
Qy    150 VD-KE-RFTTLV---QDLANAFQQEAQ-TSGKE--RLLLSAAVPAGQTYVDAGYEVDKIA 201

Db    550 qyldyvnimsydlhgawndhvghnaalydtgkdselaqw-nvygtaq 595
          | ||:||:|:||:||:|:   |||::||    :|  |   || :: |
Qy    202 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQ 248


Search completed: Fri Jun 27 16:59:22 1997
Job time : 98 secs.
```

****

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 27 16:56:14 1997; MasPar time 23.19 Seconds
572.791 Million cell updates/sec

Tabular output not generated.

Title: >US-08-663-618A-4
Description: (1-466) from US08663618A.pep
Perfect Score: 3440
Sequence: 1 MVRSVAWAGFMVLLMIPWGS.....QQSCPTGLVFSNSCKCCTWN 466

Scoring table: PAM 150
Gap 11

Searched: 89912 seqs, 28507787 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir50
1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8
13:unann9 14:unann10 15:unenc 16:unrev

Statistics: Mean 48.432; Variance 103.240; scale 0.469

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1513	44.0	383	13	A49562	cartilage glycoprote	9.27e-263
2	1456	42.3	383	16	S51327	heparin-binding glyc	1.65e-251
3	1453	42.2	381	14	S61550	BRP39 protein - mous	6.46e-251
4	1453	42.2	381	14	I48271	BRP39 protein - mous	6.46e-251
5	1374	39.9	399	5	S27879	secretory protein YM	2.39e-235
6	1355	39.4	539	14	I46470	estrogen dependent o	1.31e-231
7	1339	38.9	537	14	S57197	oviduct-specific gly	1.85e-228
8	1179	34.3	483	12	A53918	chitinase (EC 3.2.1.	4.64e-197
9	1080	31.4	504	12	A38221	chitinase (EC 3.2.1.	1.02e-177
10	1025	29.8	554	12	A56596	chitinase (EC 3.2.1.	5.30e-167
11	700	20.3	699	10	A38368	chitinase (EC 3.2.1.	2.24e-104
12	594	17.3	378	5	S51591	chitinase (EC 3.2.1.	2.53e-84
13	585	17.0	424	5	S47133	chitinase (EC 3.2.1.	1.24e-82
14	579	16.8	423	5	JQ1975	chitinase (EC 3.2.1.	1.66e-81
15	564	16.4	423	12	S51369	chitinase - fungus (	1.07e-78
16	548	15.9	427	12	JC4565	chitinase (EC 3.2.1.	1.04e-75
17	449	13.1	563	16	S60651	chitinase precursor	2.08e-57
18	407	11.8	561	9	A25090	chitinase (EC 3.2.1.	8.98e-50
19	348	10.1	799	8	PC4106	chitinase (EC 3.2.1.	3.20e-39
20	347	10.1	820	16	A40633	chitinase (EC 3.2.1.	4.81e-39
21	340	9.9	499	9	S52422	chitinase (EC 3.2.1.	8.28e-38
22	333	9.7	499	9	S04856	chitinase (EC 3.2.1.	1.41e-36
23	314	9.1	511	11	S61166	probable membrane pr	2.95e-33
24	290	8.4	610	10	JH0573	chitinase (EC 3.2.1.	4.13e-29
25	266	7.7	452	12	JC4038	47K glycoprotein - f	4.99e-25
26	266	7.7	1146	11	S07915	RF2 protein - yeast	4.99e-25
27	247	7.2	413	5	JC2135	chitinase (EC 3.2.1.	7.53e-22
28	232	6.7	124	5	S57715	chitinase (EC 3.2.1.	2.23e-19
29	172	5.0	36	14	A27682	39K whey protein - b	6.14e-10
30	172	5.0	83	13	A37954	estradiol-stimulated	6.14e-10
31	136	4.0	29	14	S57204	oviduct-specific gly	8.90e-05
32	136	4.0	29	14	S55693	oviduct-specific sia	8.90e-05
33	117	3.4	597	5	S33848	chitinase (EC 3.2.1.	2.73e-02
34	117	3.4	597	5	S32039	chitinase (EC 3.2.1.	2.73e-02
35	115	3.3	642	16	S37623	homologies with orf	4.85e-02
36	110	3.2	488	10	A41961	chitinase (EC 3.2.1.	1.98e-01
37	105	3.1	787	8	PC1232	retrovirus-related c	7.78e-01
38	105	3.1	1409	4	OFFFCP	retrovirus-related p	7.78e-01
39	103	3.0	662	11	S61193	probable membrane pr	1.33e+00
40	100	2.9	364	11	S45925	hypothetical protein	2.92e+00
41	101	2.9	562	3	KRHUEB	keratin, 56K type II	2.25e+00
42	100	2.9	616	14	S50998	transporter protein	2.92e+00
43	101	2.9	958	1	JN0102	glucan 1,4-alpha-glu	2.25e+00
44	101	2.9	1045	2	SUSEMM	serine proteinase (E	2.25e+00
45	101	2.9	1070	11	S19686	alpha-glucosidase (E	2.25e+00

ALIGNMENTS

```
RESULT    1
ENTRY             A49562     #type complete
TITLE             cartilage glycoprotein gp39 precursor - human
ALTERNATE_NAMES   39K synovial protein
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change
                    03-May-1996
ACCESSIONS        A49562; S10677; A33162
REFERENCE         A49562
   #authors       Hakala, B.E.; White, C.; Recklies, A.D.
   #journal       J. Biol. Chem. (1993) 268:25803-25810
   #title         Human cartilage gp-39, a major secretory product of articular
                    chondrocytes and synovial cells, is a mammalian member of a
                    chitinase protein family.
   #accession     A49562
      ##status          preliminary
      ##molecule_type mRNA
      ##residues       1-383 ##label HAK
      ##cross-references GB:M80927
REFERENCE         S10677
   #authors       Nyirkos, P.; Golds, E.E.
   #journal       Biochem. J. (1990) 269:265-268
   #title         Human synovial cells secrete a 39 kDa protein similar to a
                    bovine mammary protein expressed during the non-lactating
                    period.
   #cross-references MUID:90328983
   #accession     S10677
      ##molecule_type protein
      ##residues       22-40,'X',42-45 ##label NY2
KEYWORDS          cartilage; extracellular protein; glycoprotein
FEATURE
   1-21                #domain signal sequence #status predicted #label SIG\
   22-383              #product cartilage glycoprotein gp39 #status predicted
                         #label MAT
SUMMARY           #length 383  #molecular-weight 42613  #checksum 2942

  Query Match           44.0%;  Score 1513;  DB 13;  Length 383;
  Best Local Similarity  52.5%;  Pred. No. 9.27e-263;
  Matches   200;  Conservative   89;  Mismatches  83;  Indels   9;  Gaps   5;

Db       8 tgfvvlvllqccsayklvcyytswsqyregdgscfpdaldrflcthiiysfanisndhid 67
           :||:||:::   || |||||:|:|:|||:|::  :|  ||  ||||:||:|| ::| ::
Qy       8 AGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLS 67

Db      68 twewndvtlygmlntlknrnpnlktllsvggwnfgsqrfskiasntqsrrtfiksvppfl 127
```

```
          | |||| |||  :| ||: ||:|||||::|||||::|:|: : : : :|:||::|   ||
Qy      68 TTEWNDETLYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFL 127

Db     128 rthgfdgldlawlypgrr-----dkqhfttlikemkaefikeaq-pgkkqlllsaalsag 181
           |  :|||||| | ||| :     ||::|||: ::   |  ||| :|| :||||||::||
Qy     128 RKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG 187

Db     182 kvtidssydiakisqhldfisimtydfhgawrgttghhsplfrgqedaspdrfsntdyav 241
             :|::|:: ||:|:|||:::|:|||||:|   |||:|||:: ||::::    | | ||
Qy     188 QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAV 247

Db     242 gymlrlgapasklvmgiptfgrsftlass-etgvgapisgpgipgrftkeagtlayyeic 300
             |: |:|||||::|:||:|||||||||| :| |||| :|:| || ||||:| ||||||:|
Qy     248 QQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVC 307

Db     301 dflrgatvhrtlgqqvpyatkgnqwvgyddqesvkskvqylkdrqlagamvwaldlddfq 360
             :||| :|   | |||  : |||||:|| || |:|| |||:: |:||||||||||||
Qy     308 SW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFA 366

Db     361 gsfcgqdlrfpltnaikdala 381
           |  | |  |:||  :::: |:
Qy     367 GFSCNQG-RYPLIQTLRQELS 386


RESULT    2
ENTRY             S51327     #type complete
TITLE             heparin-binding glycoprotein 38K - pig
ORGANISM          #formal_name Sus scrofa domestica #common_name domestic pig
DATE              07-May-1995 #sequence_revision 07-May-1995 #text_change
                    07-May-1995
ACCESSIONS        S51327
REFERENCE         S51327
   #authors       Shackelton, L.M.; Mann, D.M.; Millis, A.J.T.
   #submission    submitted to the EMBL Data Library, January 1995
   #description   Identification of a 38kDa heparin-binding glycoprotein
                    (gp38k) in differentiating vascular smooth muscle cells as
                    a member of a group of proteins associated with tissue
                    remodeling.
   #accession     S51327
      ##status         preliminary
      ##residues       1-383 ##label SHA
      ##cross-references EMBL:Z47803
SUMMARY           #length 383  #molecular-weight 42443  #checksum 2907

  Query Match           42.3%;  Score 1456;  DB 16;  Length 383;
  Best Local Similarity  50.8%;  Pred. No. 1.65e-251;
  Matches   197;  Conservative   88;  Mismatches  94;  Indels   9;  Gaps   5;

Db       1 mglrvaqtgfvalvllqscaayklvcyytswsqyregdgscfpdaidpflcthiiysfan 60
           |   || :||: |:::   :| |||||:|:|:|||:|::  :|  :|| ||||:||:||
Qy       1 MVRSVAWAGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAG 60

Db      61 isnneidtlewndvtlydtlntlknrnpnlktllsvggwnfgsqrfskiasntqsrrtfi 120
           ::|::: | |||| ||| :| ||: ||:|||||::|||||::|:|: : : : :|:||:
Qy      61 MTNHQLSTTEWNDETLYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFV 120

Db     121 ksvppflrthgfdgldlawispgrr-----dkrhlttlvkemkaefvreal-pgterlll 174
           :|   |||  :|||||| |  || :     || ::|||| ::   | :||  :| ||||||
Qy     121 NSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLL 180

Db     175 sgavsagkvaidrgydiaqisqhldfislltydfhgawrqttghhsplfrgqgdassdrf 234
           |:||:||   :| ||::  |:|:|||::|::|||||:|   |||:|||:: | ::::
Qy     181 SAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAAS 240

Db     235 snadyavsyvlrlgapanklvmgiptfgrsftlassk-tdvgapasgpgipgrftkekgi 293
             | | ||   |: |:||:||::|:||:|||||||||  | |||||:|:| || |||| |:
Qy     241 LNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGM 300

Db     294 layyeicdflqgatvrrplgqqvpyatkgnqwvgyddqesvknkakylksrqlagamvwt 353
           |||||:|   ||| :|   | |||  : |||||:|| || | |  ||| : |:|||||:
Qy     301 LAYYEVCSW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWA 359

Db     354 ldlddfrgnfcgqnlrfpltsaikdvla 381
           |||||| |  | |  |:||  :::: |:
Qy     360 LDLDDFAGFSCNQG-RYPLIQTLRQELS 386


RESULT    3
ENTRY             S61550     #type complete
TITLE             BRP39 protein - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change
                    03-May-1996
ACCESSIONS        S61550
REFERENCE         S61550
   #authors       Morrison, B.W.
   #submission    submitted to the EMBL Data Library, November 1995
   #accession     S61550
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-381 ##label MOR
      ##cross-references EMBL:X93035
SUMMARY           #length 381  #molecular-weight 43001  #checksum 5107

  Query Match           42.2%;  Score 1453;  DB 14;  Length 381;
  Best Local Similarity  49.6%;  Pred. No. 6.46e-251;
  Matches   193;  Conservative   89;  Mismatches  96;  Indels  11;  Gaps   9;

Db       1 mgmraaltgfavlmllqscsayklvcyftswsqyregvgsflpdaiqpflcthiiysfan 60
           |    | :|| ||:::   || |||||||:|:|||:| : |||  ::| ||||:||:||
Qy       1 MVRSVAWAGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAG 60

Db      61 issdnmlstwewndesnydklnklktrntnlktllsvggwkfgekrfseiasnterrtaf 120
           :: :: ||| |||||: |: :| ||  | :|||||::||||:|:  :|::: : :: | :|
Qy      61 MT-NHQLSTTEWNDETLYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTF 119

Db     121 vrsvapflrsygfdgldlawlyp--r-lr--dkqyfstlikelnaeftkevq-pgrekll 174
           | |   ||| |:|||||| | ||  :     ||: |:||: :|   |  | | :|:|:||
Qy     120 VNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLL 179

Db     175 lsaalsagkvaidtgydiaqiaqhldfinlmtydfhgvwrqitghhsplfqgqkdtrfdr 234
           ||||::||   :|:||::  |||:|||:|||:||||| |  :|||:|||:  | ::
Qy     180 LSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA 239

Db     235 ysnvnyavqymirlgaqasklimgiptfgksftlass-enqlgapisgeglpgrftkeag 293
             ||: |||  :: |: ||||::|:||:|:|||||| : ::|||  :| | || ||||:|
Qy     240 SLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGG 299

Db     294 tlayyeicdflkgaevhrlsnekvpfatkgnqwvgyehkesvknkvgflkekklagamvw 353
           |||||:|   |||  :|: ::|||:  : |||||::  || | ||::||:| |:|||||
Qy     300 MLAYYEVCSW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358

Db     354 aldlddfqg-tcqpkeffpltnaikdala 381
           ||||||| | :|    : ||  :::: |:
Qy     359 ALDLDDFAGFSCNQGRY-PLIQTLRQELS 386


RESULT    4
ENTRY             I48271     #type complete
TITLE             BRP39 protein - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
                    02-Jul-1996
ACCESSIONS        I48271
REFERENCE         I48271
   #authors       Morrison, B.W.; Leder, P.
   #journal       Oncogene (1994) 9:3417-3426
   #title         neu and ras initiate murine mammary tumors that share genetic
                    markers generally absent in c-myc and int-2-initiated
                    tumors.
   #cross-references MUID:95060797
   #accession     I48271
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-381 ##label RES
      ##cross-references EMBL:X93035; NID:g1085065; CDS_PID:g1085066
```

```
GENETICS
   #note          gene name brp39
SUMMARY           #length 381  #molecular-weight 43001  #checksum 5107

  Query Match           42.2%;  Score 1453;  DB 14;  Length 381;
  Best Local Similarity  49.6%;  Pred. No. 6.46e-251;
  Matches   193;  Conservative   89;  Mismatches  96;  Indels  11;  Gaps   9;

Db       1 mgmraaltgfavlmllqscsayklvcyftswsqyregvgsflpdaiqpflcthiiysfan 60
           |    | :|| ||::   || |||||||:|:|||:| : |||  ::| ||||:|||:|| 
Qy       1 MVRSVAWAGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAG 60

Db      61 issdnmlstwewndesnydklnklktrntnlktllsvggwkfgekrfseiasnterrtaf 120
           :: :: ||| ||||| |: :| ||  | :|||||::|||:|:  :|::: : :: | :| 
Qy      61 MT-NHQLSTTEWNDETLYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTF 119

Db     121 vrsvapflrsygfdgldlawlyp--r-lr--dkqyfstlikelnaeftkevq-pgrekll 174
           | |   ||| |:|||||| | ||  :     ||: |:||: :|   |  | | :|:|:|| 
Qy     120 VNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLL 179

Db     175 lsaalsagkvaidtgydiaqiaqhldfinlmtydfhgvwrqitghhsplfqgqkdtrfdr 234
           ||||::||   :|:||::  |||:|||:|||:||||| |  :|||:|||:  | ::  
Qy     180 LSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA 239

Db     235 ysnvnyavqymirlgaqasklllmgiptfgksftlass-enqlgapisgeglpgrftkeag 293
           ||: |||  :: |: |||||::|:||:|:||||||| : ::||| :| | || ||||:| 
Qy     240 SLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGG 299

Db     294 tlayyeicdflkgaevhrlsnekvpfatkgnqwvgyehkesvknkvgflkekklagamvw 353
           |||||:|   |||  :|: ::|||:  : |||||::  || | ||::||:| |:|||| 
Qy     300 MLAYYEVCSW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358

Db     354 aldlddfqg-tcqpkeffpltnaikdala 381
           ||||||| | :|    : ||  :::: |: 
Qy     359 ALDLDDFAGFSCNQGRY-PLIQTLRQELS 386


RESULT    5
ENTRY             S27879     #type complete
TITLE             secretory protein YM-1 precursor - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
                    03-May-1996
ACCESSIONS        S27879
REFERENCE         S27879
   #authors       Chang, N.C.A.; Liu, C.H.; Chang, A.C.
   #submission    submitted to the EMBL Data Library, June 1992
   #description   Molecular characterization of a secretory protein (YM-1)
                    transiently expressed by activated murine peritoneal
                    macrophages.
   #accession     S27879
      ##molecule_type mRNA
      ##residues      1-399 ##label CHA
      ##cross-references EMBL:M94584
CLASSIFICATION    #superfamily Streptomyces chitinase chi40
FEATURE
   1-21               #domain signal sequence #status predicted #label SIG\
   22-399             #product secretory protein YM-1 #status predicted #label
                        MAT
SUMMARY           #length 399  #molecular-weight 44528  #checksum 3208

  Query Match           39.9%;  Score 1374;  DB 5;  Length 399;
  Best Local Similarity  46.2%;  Pred. No. 2.39e-235;
  Matches   180;  Conservative   95;  Mismatches 112;  Indels   3;  Gaps   3;

Db       1 maklilvtglaillnvqlgssyqlmcyytswakdrpiegsfkpgnidpclcthliyafag 60
           | : :  :|: :|| :  ||:  |:||:|:||  |  |: | | ::|| ||||||||||| 
Qy       1 MVRSVAWAGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAG 60

Db      61 mqnneitytheqdlrdyealnglkdkkntelktllaiggwkfgpapfsamvstpqnrqif 120
           | |:::: |   |   |: :|||| |:|  |||||||||:|:   |: ||:|: ||| | 
Qy      61 MTNHQLSTTEWNDETLYQEFNGLK-KMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTF 119

Db     121 iqsvirflrqynfdglnldwqypgsrgsppkdkhlfsvlvkemrkafeeesvekdiprll 180
           : | ||||| |:||||:|||:||||:|||: ||  |: || :: :||::|:      ||| 
Qy     120 VNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLL 179

Db     181 ltstgagiidviksgtrslnclslldyiqvmtydlhdpkdgytgensplykspydigksa 240
           |::: ::    : :|    :    ||:: :|:||:| : :  || ||||||   : | :|  
Qy     180 LSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA 239

Db     241 dlnvdsiisywkdhgaaseklivgfpayghtfilsdpsktgigaptistgppgkytdesg 300
           ||||: :  | : |::: |||:| |:||::| |: :| | :|||: ::| || :| |:| 
Qy     240 SLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGG 299

Db     301 llayyevctflnegatevwdapqevpyayqgnewvgydnvrsfklkaqwlkdnnlggavv 360
           :|||||||:  : |||  :   | ||| :: |:|||:|:| ||| |   ||:: ||||:| 
Qy     300 MLAYYEVCSW-K-GATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMV 357

Db     361 wpldmddfsgsfchqrhfpltstlkgdlni 390
           |:|||:|||:|  |:| ::||  ||: :|:: 
Qy     358 WALDLDDFAGFSCNQGRYPLIQTLRQELSL 387


RESULT    6
ENTRY             I46470     #type complete
TITLE             estrogen dependent oviduct protein precursor - sheep
ORGANISM          #formal_name Ovis orientalis aries, Ovis ammon aries
                    #common_name domestic sheep
DATE              16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change
                    16-Aug-1996
ACCESSIONS        I46470
REFERENCE         I46470
   #authors       DeSouza, M.M.; Murray, M.K.
   #journal       Endocrinology (1995) 136:2485-2496
   #title         An estrogen-dependent secretory protein, which shares
                    identity with chitinases, is expressed in a temporally and
                    regionally specific manner in the sheep oviduct at the time
                    of fertilization and embryo development.
   #cross-references MUID:95269691
   #accession     I46470
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-539 ##label DES
      ##cross-references EMBL:U16719; NID:g885600; CDS_PID:g885601
SUMMARY           #length 539  #molecular-weight 59535  #checksum 2120

  Query Match           39.4%;  Score 1355;  DB 14;  Length 539;
  Best Local Similarity  48.6%;  Pred. No. 1.31e-231;
  Matches   188;  Conservative   86;  Mismatches 106;  Indels   7;  Gaps   5;

Db       1 mgklllwvglllmlkhhdgaahklvcyftnwafsrpgsasilprdldpflcthlvfafas 60
           | : : | |::::|:   |:| ||||||||||  | | | :||:|||| |||||::|||: 
Qy       1 MVRSVAWAGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAG 60

Db      61 mnnnqivpkdpldekilypefnklkernrglktllsvggwnfgtsrftkmlstfsnrerf 120
           | |:|:   :  ::  || ||| ||  |  ||||||::||||||:| :|| |::| :||: | 
Qy      61 MTNHQLSTTE-WNDETLYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTF 119

Db     121 vksviallrthgfdgldlfflypglrgspardrwtfvflleellqafkneaqltmrprll 180
           |:| | :||  :|||||   ||| :|||| |:  |  |:::|  ||  ||| : : |||  
Qy     120 VNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLL 179

Db     181 lsaavsgdphviqkaydarllgrlldfisvlsydlhgswekvtghnsplfslpgdpkssa 240
           ||||||::    :: :|:   ::: |||:::::||:|||||||||||||:    :: ::| 
Qy     180 LSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA 239

Db     241 -yam--sy--wrqlgvppekllmglptygrtfhllrasqnelgagaagpaspgkytkqag 295
            :  :   | | | |: |||:|:|||||:| |  :|:  :|| |:|:::|| :||::| 
Qy     240 SLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGG 299

Db     296 flayyevcsfvqrakkrwindqyvpyafkgkewvgyddaisfgykaffikrehfggamvw 355
           |||||||    | |::| || ||| |: ::|||:||  ||  |  ::|:  :||||||  
Qy     300 MLAYYEVCSW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358

Db     356 tldlddfrgnfcgtgpfplahtlnnll 382
```

```
          :|||||| |  |  | :|| :||   |
Qy    359 ALDLDDFAGFSCNQGRYPLIQTLRQEL 385


RESULT    7
ENTRY             S57197     #type fragment
TITLE             oviduct-specific glycoprotein 95K precursor - bovine
                    (fragment)
ORGANISM          #formal_name Bos primigenius taurus #common_name cattle
DATE              28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
                    03-May-1996
ACCESSIONS        S57197
REFERENCE         S57197
   #authors       Sendai, Y.; Abe, H.; Kikuchi, M.; Satoh, T.; Hoshi, H.
   #journal       Biol. Reprod. (1994) 50:927-934
   #title         Purification and molecular cloning of bovine oviduct-specific
                    glycoprotein.
   #accession     S57197
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-537 ##label SEN
      ##cross-references EMBL:D16639
SUMMARY           #length 537  #checksum 1695

  Query Match           38.9%;  Score 1339;  DB 14;  Length 537;
  Best Local Similarity  50.1%;  Pred. No. 1.85e-228;
  Matches   190;  Conservative   81;  Mismatches 101;  Indels   7;  Gaps   6;

Db      6 glllvlkhhdgaahklvcyftnwafsrpgpasilprdldpflcthlvfafasmsnnqivp 65
          |::::|:   |:| ||||||||||  | | | :||:|||| |||||::|||:|:|:|:
Qy      9 GFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLST 68

Db     66 kdpqdekilypefnklkernrglktllsiggwnfgtvrfttmlstfsnrerfvssviall 125
           :  ||  || ||| ||  |  |||||:||||||:| :|| |::| :||: ||:| | :|
Qy     69 TEWNDET-LYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFL 127

Db    126 rthgfdgldlfflypglrgspardrwtfvflleellqafkneaqltmrprlllsaavsgd 185
          |  :||||||   ||| :|||| |:  |  |:::|  ||  ||| : : |||||||::
Qy    128 RKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG 187

Db    186 phvvqkayearllgrlldfisvlsydlhgswekvtghnsplfs-l--pgdpks-sa-yam 240
            |: :||   ::: |||:::::||:|||||||||||||||:     :| : | :   |:
Qy    188 QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAV 247

Db    241 nywrqlgvppekllmglptygrtfhllkasqnelraqavgpaspgkytkqagflayyeic 300
            | | | |: ||::|:|||||:| |  :|:  : | | |:::|| :||::| |||||:|
Qy    248 QQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVC 307

Db    301 cfvrrakkrwindqyvpyafkgkewvgyddaisfgykaffikrehfggamvwtldlddfr 360
            : | |::| || ||| |: ::|||:||  ||  |  ::|:  :|||||:||||||
Qy    308 SW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFA 366

Db    361 gyfcgtgpfplvhtlnnll 379
          |: |  | :||::||   |
Qy    367 GFSCNQGRYPLIQTLRQEL 385


RESULT    8
ENTRY             A53918     #type complete
TITLE             chitinase (EC 3.2.1.14) precursor - braconid wasp (Chelonus
                    sp.)
ORGANISM          #formal_name Chelonus sp.
DATE              28-Jul-1995 #sequence_revision 28-Jul-1995 #text_change
                    03-May-1996
ACCESSIONS        A53918
REFERENCE         A53918
   #authors       Krishnan, A.; Nair, P.N.; Jones, D.
   #journal       J. Biol. Chem. (1994) 269:20971-20976
   #title         Isolation, cloning, and characterization of new chitinase
                    stored in active form in chitin-lined venom reservoir.
   #accession     A53918
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-483 ##label KRI
      ##cross-references GB:U10422
KEYWORDS          glycosidase; hydrolase
SUMMARY           #length 483  #molecular-weight 52013  #checksum 5938

  Query Match           34.3%;  Score 1179;  DB 12;  Length 483;
  Best Local Similarity  42.0%;  Pred. No. 4.64e-197;
  Matches   158;  Conservative  101;  Mismatches 100;  Indels  17;  Gaps  11;

Db     19 aspnkvvcyfgawsvyrqgngkfdingidptlcthliysfvgvngkdvkvldpwsdlpgn 78
          :|: |:|||  |: ||||:::|  : :||:|||||||:| |:   ::   : |:|
Qy     19 GSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTE-WND--ET 75

Db     79 ldgfgkftslrkknpsvkimvavggwnagsvpfsqmasdqatreafaqnvvkflqqyqfd 138
          |  :  | :|:|:|| :| ::|:|||| ::  |::| :    |::|  : ::||: | ||
Qy     76 L--YQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFD 133

Db    139 gfdidweypaqrggspadvknmvklckalkkafvq--h----d-yilsaavaapetsask 191
          |:|:|||||: :|::: |      |   | :|| |  :    :  :|||||:| :|
Qy    134 GLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDA 193

Db    192 sydiaemsqyldfinlmtydfhgpwdghtgmhappsasshdsgnelklnvkaavkywlqn 251
          :|::  ::| |||:|||:|||||:|:  || ::|      :||    ||| |||  |||:
Qy    194 GYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQK 253

Db    252 gvpkeklvvgvpaygksftlsnpsnkglgapvsgagtagpytgengllgyneicemqkag 311
          | |  ||::|:|:||:||||:::|:  :||| :|:||:||:| | |:|:| |:|   |::
Qy    254 GTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGA 312

Db    312 dwevvqdnekgvpyavkgnqwvsfddlaaikakaqfikqeglggamvwsietddfkgl-c 370
            : :|| :| |||  : ||||:|||: ::|:|  ::|| |||||||||::: ||| |: |
Qy    313 TKQRIQD-QK-VPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSC 370

Db    371 ge-kypvlkalnsvlg 385
           : :||:: :|   |:
Qy    371 NQGRYPLIQTLRQELS 386


RESULT    9
ENTRY             A38221     #type complete
TITLE             chitinase (EC 3.2.1.14) MF1 - nematode (Brugia malayi)
ORGANISM          #formal_name Brugia malayi
DATE              04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
                    03-May-1996
ACCESSIONS        A38221
REFERENCE         A38221
   #authors       Fuhrman, J.A.; Lane, W.S.; Smith, R.F.; Piessens, W.F.;
                    Perler, F.B.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1992) 89:1548-1552
   #title         Transmission-blocking antibodies recognize microfilarial
                    chitinase in brugian lymphatic filariasis.
   #cross-references MUID:92179220
   #accession     A38221
      ##status         preliminary; not compared with conceptual translation
      ##molecule_type nucleic acid; protein
      ##residues       1-504 ##label FUH
      ##cross-references NCBIP:85345
      ##note          sequence extracted from NCBI backbone
KEYWORDS          glycosidase; hydrolase
SUMMARY           #length 504  #molecular-weight 55971  #checksum 7343

  Query Match           31.4%;  Score 1080;  DB 12;  Length 504;
  Best Local Similarity  42.5%;  Pred. No. 1.02e-177;
  Matches   157;  Conservative   95;  Mismatches 102;  Indels  15;  Gaps  11;

Db     27 cyytnwaqyrdgegkflpgnipnglcthilyafakvdelgdskpfewndedtewskgmys 86
          ||:|||||||:|||::||| ::  :||||::|||| : :   |   |||||  |     :|
Qy     26 CYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTT-EWNDE-T-----LYQ 78

Db     87 avtklretnpglkvllsyggynfgsaiftgiaksaqkterfiksaiaflrknnfdgfdld 146
              |:  || || ||: || ||::  || :  :| : : |::||| |||| :|||:|||
Qy     79 EFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 138
```

```
Db    147 weyp-v-gv-a-e-eh-aklveamktafveeaktsgkqrllltaavsagkgtidgsynve 200
          ||||   |  | : |: : ||: :  || :|| ||||:||||:|||:||   :|::|:|:
Qy    139 WEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVD 198

Db    201 slgknfdllflmsydlhgsweknvdlhgklhptkgevsgigifntefaadywaskgmpke 260
           :: |:|:: ||:||:|||||    :: |     | :: : :| : | : |  || |
Qy    199 KIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPAS 258

Db    261 kiiigipmyaqgwtldnpsetaigaaasrpssasktnpaggtasyweickylkeggketv 320
          |:|:|:| |::: || ::|:| :||:|: :::::     ||  :| |:|   | : |: :
Qy    259 KLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRI 317

Db    321 hqegvgaymvkgdqwygydneetirikmkwlkekgyggafiwaldfddftgkscgkgpyp 380
          ::: | :|: : :|| |:|: |::: |:  ||:|| ||| :||||:|||:| ||  | ||
Qy    318 QDQKV-PYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYP 376

Db    381 llnaissel 389
          |: ::  ||
Qy    377 LIQTLRQEL 385


RESULT   10
ENTRY            A56596     #type complete
TITLE            chitinase (EC 3.2.1.14) - tobacco hornworm
ORGANISM         #formal_name Manduca sexta #common_name tobacco hornworm
DATE             11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change
                   03-May-1996
ACCESSIONS       A56596
REFERENCE        A56596
   #authors      Kramer, K.J.; Corpuz, L.; Choi, H.K.; Muthukrishnan, S.
   #journal      Insect Biochem. Mol. Biol. (1993) 23:691-701
   #title        Sequence of a cDNA and expression of the gene encoding
                   epidermal and gut chitinases of Manduca sexta.
   #cross-references MUID:93357793
   #accession    A56596
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-554 ##label KRA
      ##cross-references NCBIN:136417; NCBIP:136418
      ##experimental_source larvae
      ##note          sequence extracted from NCBI backbone
KEYWORDS         glycosidase; hydrolase
SUMMARY          #length 554  #molecular-weight 62203  #checksum 4659

  Query Match           29.8%;  Score 1025;  DB 12;  Length 554;
  Best Local Similarity  38.5%;  Pred. No. 5.30e-167;
  Matches   161;  Conservative  106;  Mismatches 126;  Indels  25;  Gaps  19;

Db     17 vqsdsrarivcyfsnwavyrpgvgrygiedipvekcthiiysfigvtegnsevliidpel 76
          :   | |::|||||:||| || | :|:   |:    |||:||:| |:|: : : |    |
Qy     16 IPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTN-H-Q-LST-TEW 71

Db     77 dvdkngfrnftslrsshpsvkfmvavggwaegsskyshmvaqkstrmsfirsvvsflkky 136
          : |   :::| :|:  :| :| ::|:|||  :: |:: |||  : | :|: | : ||:||
Qy     72 N-DETLYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKY 130

Db    137 dfdgldldweypgaadrggsfsdkdkflylvqelrraf---irv-gkgwel-taavplan 191
           ||||||||||||:  :|::  ||::|  |||:|  ||    :  || : | :|||| :
Qy    131 SFDGLDLDWEYPGS--QGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQ 188

Db    192 frlmegyhvpelcqeldaihvmsydlrgnwagfadvhsplykrphdqwayeklnvndglh 251
          :  || |  : |:|| :::|:||::|:|   :  :|||||  :  |   |||: :::
Qy    189 TYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQ 248

Db    252 lweekgcpsnklvvgipfygrsftlsagnnnyglgtfinkeagggdpapytnatgfwayy 311
           | :|| |::||::|:| |||||||::::   :|:     :|:| |:|:|:  |  |||
Qy    249 QWLQKGTPASKLILGMPTYGRSFTLASSSDT-RVGA--P-ATGSGTPGPFTKEGGMLAYY 304

Db    312 eictevdkddsgwtkkwdeqgkcpyaykgtqwvgyedprsveikmnwikqkgylgamtwa 371
          |:|:   |  :  || : :: | || ::  ||||::|  |   |:: :|||| ||| ||
Qy    305 EVCS--WKG-A--TKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWA 359

Db    372 idmddfqgl-cge-knplikilhkhms-sytvppphtenttptpewarppstpsdpse 426
          :|:|||| |: | : : |||  |:  :| :| :|:   |   | |  :  |   : |::
Qy    360 LDLDDFAGFSCNQGRYPLIQTLRQELSLPY-LPSGTPELEVPKPGQPSEPEHGPSPGQ 416


RESULT   11
ENTRY            A38368     #type complete
TITLE            chitinase (EC 3.2.1.14) precursor - Bacillus circulans
ORGANISM         #formal_name Bacillus circulans
DATE             28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change
                   03-May-1996
ACCESSIONS       A38368
REFERENCE        A38368
   #authors      Watanabe, T.; Suzuki, K.; Oyanagi, W.; Ohnishi, K.; Tanaka,
                   H.
   #journal      J. Biol. Chem. (1990) 265:15659-15665
   #title        Gene cloning of chitinase A1 from Bacillus circulans WL-12
                   revealed its evolutionary relationship to Serratia
                   chitinase and to the type III homology units of
                   fibronectin.
   #cross-references MUID:90368776
   #accession    A38368
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-699 ##label WAT
      ##cross-references GB:J05599
CLASSIFICATION   #superfamily fibronectin type III repeat homology
KEYWORDS         glycosidase; hydrolase
SUMMARY          #length 699  #molecular-weight 73677  #checksum 6757

  Query Match           20.3%;  Score 700;  DB 10;  Length 699;
  Best Local Similarity  40.4%;  Pred. No. 2.24e-104;
  Matches   120;  Conservative   72;  Mismatches  85;  Indels  20;  Gaps  18;

Db    144 qlnklkqtnpnlktiisvggwtwsn-rfsdvaataatrevfansavdflrkynfdgvdld 202
          ::| ||  ||:|||:::::|||  |  :|:|: |||  |: | |||: |||||:|||:|||
Qy     79 EFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 138

Db    203 weypvsggldgnskrpedkqnytlllskirekldaagavdgk-kylltiasgasatya-a 260
          ||||  |: :| |  : ||: :| |:  : : : ::  :   || : ||: |  |: ||  |
Qy    139 WEYP--GS-QG-S-PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDA 193

Db    261 ntelakiaaivdwinimtydfngawqkisahnaplnydpaasaagvpdantfnvaagaqg 320
            |: |||  :| :|:|:|||:|:|:|:::||:|| |      :| : | ::|| |: |
Qy    194 GYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPL-YKRQE-ESGAA-A-SLNVDAAVQQ 249

Db    321 hldagvpaaklvlgvpfygrgwd-gcaqagn-gqyqtctggssvgtweagsfdfydlean 378
          |: | ||:||:||:| |||:   : :      |   | :| ::  | |:| : :|::  :
Qy    250 WLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEV-CS 308

Db    379 yinkngytrywndtakvpylynasnkrfisyddaesvgyktayikskglggamfwel 435
            | | |: : :  |||::   | : :::|| ||   | :|:| |||||| | |
Qy    309 W--K-GATKQRIQDQKVPYIFRD-NQW-VGFDDVESFKTKVSYLKQKGLGGAMVWAL 360


RESULT   12
ENTRY            S51591     #type complete
TITLE            chitinase (EC 3.2.1.14)/lysozyme (EC 3.2.1.17) PZ precursor,
                   pathogenesis-related - common tobacco
ORGANISM         #formal_name Nicotiana tabacum #common_name common tobacco
DATE             15-Jul-1995 #sequence_revision 08-Sep-1995 #text_change
                   03-May-1996
ACCESSIONS       S51591; S51632; S43119
REFERENCE        S51591
   #authors      Heitz, T.; Segond, S.; Kauffmann, S.; Geoffroy, P.; Prasad,
                   V.; Brunner, F.; Fritig, B.; Legrand, M.
   #journal      Mol. Gen. Genet. (1994) 245:246-254
   #title        Molecular characterization of a novel tobacco
                   pathogenesis-related (PR) protein: a new plant
                   chitinase/lysozyme.
   #accession    S51591
      ##molecule_type mRNA
      ##residues      1-378 ##label HEI
      ##cross-references EMBL:X78325
```

```
     ##experimental_source cv. Samsun NN
   #accession    S51632
     ##molecule_type protein
     ##residues      31-36;87-112;252-275;282-305;337-371 ##label HEW
CLASSIFICATION   #superfamily Streptomyces chitinase chi40
KEYWORDS         glycosidase; hydrolase
FEATURE
   1-25               #domain signal sequence #status predicted #label SIG\
   26-371             #product chitinase/lysozyme PZ #status predicted #label
                        MAT\
   371-378            #domain carboxyl-terminal propeptide #status
                        experimental #label PRO
SUMMARY          #length 378  #molecular-weight 42019  #checksum 6689

  Query Match           17.3%;  Score 594;  DB 5;  Length 378;
  Best Local Similarity  32.7%;  Pred. No. 2.53e-84;
  Matches   106;  Conservative   87;  Mismatches 108;  Indels  23;  Gaps  18;

Db      43 nnidstlfthlfcafadlnpqsnqliispenqds-fsqftstvqrknpsvktflsiaggr 101
           :::|::| |||: ||| :   ::||  :  |::: : :| : : ::|| :||:|:| ||:
Qy      44 KDLDPSLCTHLIYAFAGMT--NHQLSTTEWNDETLYQEFNG-LKKMNPKLKTLLAI-GGW 99

Db     102 a-dttaygimarqpnsrksfidssirlarqfgfhgldldweyp--l-s-atdmtnlgill 156
             |  :  |   :|:| :|::|:||: | ::| |||||||||    | | |:  :  |:
Qy     100 NFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLV 159

Db     157 newrtainmearnsgraallltaavsysprvnglnypvesvarnlnwinlmaydfygpnw 216
            :   |:  ||: ||:  |||:|||: :      | |: :|:||: :|||||| |: |
Qy     160 QDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGS-W 218

Db     217 spsqtnshaqlfdpvnh--isgs---dg-inawiqagvptkklvlgipfygyawrlvnpn 270
              | :: |:   :    ::|   |: :  |:| | |: ||:||:| || :  | :::
Qy     219 EKV-TGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSS 277

Db     271 ihdlrapaagksnvga-vddgsm-tynrirdyivqsrattvynativgdycysgsnwisy 328
            : |||:| :  |:   :|:| :|  :      ::   : :   |  | :  : |:::
Qy     278 DTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQR-IQDQK-VP-YIFRDNQWVGF 334

Db     329 ddtqsvrnkvnyvkgrgllgyfaw 352
           || :| : ||:|:| :|| |   |
Qy     335 DDVESFKTKVSYLKQKGLGGAMVW 358


RESULT   13
ENTRY            S47133     #type complete
TITLE            chitinase (EC 3.2.1.14) - fungus (Trichoderma harzianum)
ORGANISM         #formal_name Trichoderma harzianum
DATE             23-Nov-1994 #sequence_revision 03-Aug-1995 #text_change
                   10-May-1996
ACCESSIONS       S47133
REFERENCE        S47133
   #authors      Carsolio, C.; Gutierrez, A.; Jimenez, B.; van Montagu, M.;
                   Herrera-Estrella, A.
   #submission   submitted to the EMBL Data Library, June 1994
   #description  Characterization of ech-42, a Trichoderma harzianum
                   endochitinase gene expressed during mycoparasitism.
   #accession    S47133
     ##status         preliminary
     ##molecule_type DNA
     ##residues      1-424 ##label CAR
     ##cross-references EMBL:X79381
GENETICS
   #introns      48/2; 81/2; 98/1
CLASSIFICATION   #superfamily Streptomyces chitinase chi40
KEYWORDS         glycosidase; hydrolase
SUMMARY          #length 424  #molecular-weight 46298  #checksum 8773

  Query Match           17.0%;  Score 585;  DB 5;  Length 424;
  Best Local Similarity  36.4%;  Pred. No. 1.24e-82;
  Matches   103;  Conservative   68;  Mismatches  95;  Indels  17;  Gaps  13;

Db     112 qlfk-lkkanrnlkvmlsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidv 169
           | |: ||| | :|| :|:|||| || :|    :|  ||  | ::|| |::  :|||:|:
Qy      78 QEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 137

Db     170 dweyp-addtqatnmvlllkeirsqldayaaqy-apg-yhfllsiaapagpehy-sflhm 225
           ||||| :: : | ::  :   ::   :|:  :  ::|  ::||| | |||     :   :
Qy     138 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 197

Db     226 sdlgqvldyvnlmaydyagswssysghdanlfanpsnpnssp-yntdqaikdyikggvpa 284
             ::| ||:|||||||: |||   :||:: |:    :: :::  | | |: : :  | ||
Qy     198 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 257

Db     285 skivlgmpiygraf--estgg--igqtysgigs-gsw--engiwdy-kvlpkagatvqyd 336
           ||::|||| |||:|   |::   :|   :| |: |:   | |:  |  | :  ||| |
Qy     258 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQ-- 315

Db     337 svaqayysydpsskelisfdtpdmintkvsylknlglggsmfw 379
            : :   :|    :: ::||  : ::|||||||  ||||:| |
Qy     316 RIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT   14
ENTRY            JQ1975     #type complete
TITLE            chitinase (EC 3.2.1.14) 1 precursor - imperfect fungus
                   (Aphanocladium album)
ORGANISM         #formal_name Aphanocladium album
DATE             03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change
                   03-May-1996
ACCESSIONS       JQ1975
REFERENCE        JQ1975
   #authors      Blaiseau, P.L.; Lafay, J.F.
   #journal      Gene (1992) 120:243-248
   #title        Primary structure of a chitinase-encoding gene (chi1) from
                   the filamentous fungus Aphanocladium album: Similarity to
                   bacterial chitinases.
   #accession    JQ1975
     ##molecule_type mRNA
     ##residues      1-423 ##label BLA
     ##cross-references GB:X64104
COMMENT     This enzyme is essential for the degradation of insoluble chitin.
GENETICS
   #gene         chi1
   #introns      47/2; 80/2; 97/1
CLASSIFICATION   #superfamily Streptomyces chitinase chi40
KEYWORDS         glycosidase; hydrolase
FEATURE
   1-34               #domain signal sequence #status predicted #label SIG\
   35-423             #product chitinase 1 #status predicted #label MAT
SUMMARY          #length 423  #molecular-weight 46072  #checksum 8802

  Query Match           16.8%;  Score 579;  DB 5;  Length 423;
  Best Local Similarity  36.6%;  Pred. No. 1.66e-81;
  Matches   102;  Conservative   64;  Mismatches  95;  Indels  18;  Gaps  16;

Db     115 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 173
           ||| | ::| :|:|||| || :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy      83 LKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 142

Db     174 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 228
            :::: | ::  :   |:   ::: :  : |  ::||| | ||| : |    :  : ::
Qy     143 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 201

Db     229 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 287
             ||::|||||: ||| : |||:: :|    ::: |::  | | |||  :  | ||:|::
Qy     202 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 261

Db     288 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 339
           |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy     262 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 321

Db     340 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 378
             | :    :: ::||    : |||| || |||||:| |
Qy     322 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358
```

```
RESULT   15
ENTRY            S51369     #type complete
TITLE            chitinase - fungus (Trichoderma harzianum)
ORGANISM         #formal_name Trichoderma harzianum
DATE             15-Jul-1995 #sequence_revision 19-Apr-1996 #text_change
                   03-May-1996
ACCESSIONS       S51369
REFERENCE        S51369
   #authors      Garcia, I.; Lora, J.M.; de la Cruz, J.; Benitez, T.; Llobell,
                   A.; Pintor-Toro, J.A.
   #journal      Curr. Genet. (1994) 27:83-89
   #title        Cloning and characterization of a chitinase (CHIT42) cDNA
                   from the mycoparasitic fungus Trichoderma harzianum.
   #accession    S51369
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-423 ##label GAR
SUMMARY          #length 423  #molecular-weight 46056  #checksum 2000

  Query Match           16.4%;  Score 564;  DB 12;  Length 423;
  Best Local Similarity  36.0%;  Pred. No. 1.07e-78;
  Matches   102;  Conservative   70;  Mismatches  94;  Indels  17;  Gaps  13;

Db    111 qlfk-vkkanrglkvllsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidi 168
          | |: :|| |  || ||:|||  || :|    :|  ||  | ::|| |::  :|||:|:
Qy     78 QEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 137

Db    169 dweyp-adatqasnmilllkevrsqrdayaaqy-apg-yhflltiaapagkdny-sklrl 224
          ||||| :::: | ::  :   |:   :|:  :  ::|  ::||: | |||    :   :
Qy    138 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 197

Db    225 adlgqvldyinlmaydyagsfspltghdanlfnnpsnpnatp-fntdsavkdyinggvpa 283
            ::| ||::|||||: ||   :|||:: |::   :: |:: :| |:|| : :  | ||
Qy    198 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 257

Db    284 nkivlgmpiygrsf--qntag--igqtyngvgs-gsw--eagiwdykalpk-agatvqyd 335
          :|::|||| |||||   :::   :|    | |: |:   |:|:  |  :    ||| |
Qy    258 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRI 317

Db    336 svakgyysynsatkelisfdtpdmintkvaylkslglggsmfw 378
              | | :    :: ::||  : ::|||:|||  ||||:| |
Qy    318 QDQKVPYIFRD--NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


Search completed: Fri Jun 27 16:57:26 1997
Job time : 72 secs.
```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 27 16:54:21 1997; MasPar time 14.84 Seconds
666.076 Million cell updates/sec

Tabular output not generated.

Title: >US-08-663-618A-4
Description: (1-466) from US08663618A.pep
Perfect Score: 3440
Sequence: 1 MVRSVAWAGFMVLLMIPWGS.....QQSCPTGLVFSNSCKCCTWN 466

Scoring table: PAM 150
Gap 11

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot34
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11

Statistics: Mean 50.039; Variance 84.025; scale 0.596

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1513	44.0	383	4	GP39_HUMAN	CARTILAGE GLYCOPROTEI	0.00e+00
2	1080	31.4	504	2	CHIT_BRUMA	ENDOCHITINASE PRECURS	4.93e-225
3	1025	29.8	554	2	CHIT_MANSE	ENDOCHITINASE PRECURS	1.84e-211
4	700	20.3	699	2	CHI1_BACCI	CHITINASE A1 PRECURSO	3.64e-132
5	579	16.8	423	2	CHI1_APHAL	CHITINASE 1 PRECURSOR	3.08e-103
6	564	16.4	423	2	CHI4_TRIHA	42 KD ENDOCHITINASE P	1.10e-99
7	548	15.9	427	2	CHI1_COCIM	ENDOCHITINASE 1 PRECU	6.62e-96
8	440	12.8	563	2	CHIA_SERMA	CHITINASE A PRECURSOR	1.04e-70
9	398	11.6	551	2	CHIT_NPVAC	PROBABLE ENDOCHITINAS	4.28e-61
10	348	10.1	820	2	CHIA_ALTSO	CHITINASE A PRECURSOR	7.81e-50
11	333	9.7	499	2	CHIB_SERMA	CHITINASE B PRECURSOR	1.68e-46
12	304	8.8	619	2	CHIT_STRLI	CHITINASE C PRECURSOR	3.91e-40
13	290	8.4	610	2	CHIT_STRPL	CHITINASE 63 PRECURSO	4.24e-37
14	266	7.7	1146	5	KTXA_KLULA	KILLER TOXIN ALPHA AN	5.77e-32
15	172	5.0	36	4	GP39_BOVIN	CARTILAGE GLYCOPROTEI	4.73e-13
16	172	5.0	83	3	EDOS_PAPAN	ESTRADIOL-DEPENDENT O	4.73e-13
17	154	4.5	541	11	YKC5_CAEEL	HYPOTHETICAL 55.6 KD	9.26e-10
18	117	3.4	597	2	CHIX_STROI	EXOCHITINASE 1 PRECUR	1.46e-03
19	110	3.2	497	2	CHID_BACCI	CHITINASE D PRECURSOR	1.66e-02
20	102	3.0	452	4	HEMF_RHOSH	COPROPORPHYRINOGEN II	2.35e-01
21	99	2.9	289	11	YFK2_YEAST	HYPOTHETICAL 31.9 KD	6.08e-01
22	100	2.9	364	11	YBR5_YEAST	HYPOTHETICAL 40.9 KD	4.44e-01
23	100	2.9	398	9	SPEB_STRPY	EXOTOXIN TYPE B PRECU	4.44e-01
24	101	2.9	958	1	AMYG_SCHOC	GLUCOAMYLASE 1 PRECUR	3.23e-01
25	101	2.9	1045	7	PRTT_SERMA	EXTRACELLULAR SERINE	3.23e-01
26	101	2.9	1070	1	AGLU_CANTS	ALPHA-GLUCOSIDASE PRE	3.23e-01
27	96	2.8	297	5	HMX1_HUMAN	HOMEOBOX PROTEIN MSX-	1.54e+00
28	95	2.8	353	11	YQHT_BACSU	PUTATIVE PEPTIDASE IN	2.08e+00
29	98	2.8	422	10	UMUC_ECOLI	UMUC PROTEIN.	8.30e-01
30	98	2.8	492	9	SYK_THETH	LYSYL-TRNA SYNTHETASE	8.30e-01
31	96	2.8	554	5	HYES_RAT	SOLUBLE EPOXIDE HYDRO	1.54e+00
32	95	2.8	554	5	HYES_MOUSE	SOLUBLE EPOXIDE HYDRO	2.08e+00
33	97	2.8	563	5	K2CB_HUMAN	KERATIN, TYPE II CYTO	1.13e+00
34	97	2.8	563	5	K2CA_HUMAN	KERATIN, TYPE II CYTO	1.13e+00
35	97	2.8	563	5	K2CC_HUMAN	KERATIN, TYPE II CYTO	1.13e+00
36	95	2.8	563	5	K2CE_HUMAN	KERATIN, TYPE II CYTO	2.08e+00
37	95	2.8	563	5	K2CF_HUMAN	KERATIN, TYPE II CYTO	2.08e+00
38	98	2.8	1045	7	PRTS_SERMA	EXTRACELLULAR SERINE	8.30e-01
39	95	2.8	1333	10	VGR1_MOUSE	VASCULAR ENDOTHELIAL	2.08e+00
40	95	2.8	1409	2	COPI_DROME	COPIA PROTEIN.	2.08e+00
41	93	2.7	99	10	Y150_NPVAC	HYPOTHETICAL 11.2 KD	3.78e+00
42	93	2.7	267	3	EBAG_FLASP	ENDO-BETA-N-ACETYLGLU	3.78e+00
43	93	2.7	295	11	YHO8_YEAST	HYPOTHETICAL 35.1 KD	3.78e+00
44	94	2.7	344	4	GCH2_PICGU	GTP CYCLOHYDROLASE II	2.81e+00
45	94	2.7	993	11	YAJ1_SCHPO	PUTATIVE FAMILY 31 GL	2.81e+00

ALIGNMENTS

RESULT 1
ID GP39_HUMAN STANDARD; PRT; 383 AA.
AC P36222; P30923;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE CARTILAGE GLYCOPROTEIN-39 PRECURSOR (GP-39) (39 KD SYNOVIAL PROTEIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=CARTILAGE;
RX MEDLINE; 94064658.
RA HAKALA B.E., WHITE C., RECKLIES A.D.;
RL J. BIOL. CHEM. 268:25803-25810(1993).
RN [2]
RP SEQUENCE OF 22-45.
RX MEDLINE; 90328983.
RA NYIRKOS P., GOLDS E.E.;
RL BIOCHEM. J. 269:265-268(1990).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE CAPACITY OF CELLS TO
CC RESPOND TO AND COPE WITH CHANGES IN THEIR ENVIRONMENT.
CC -!- TISSUE SPECIFICITY: PRESENT IN ARTICULAR CHONDROCYTES, SYNOVIAL
CC CELLS AS WELL AS IN LIVER. UNDETECTABLE IN MUSCLE TISSUES, LUNG,
CC PANCREAS, MONONUCLEAR CELLS, OR FIBROBLASTS.
CC -!- PTM: GLYCOSYLATED.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
DR EMBL; M80927; G348912; -.
DR PIR; S10677; S10677.
DR PIR; A33162; A33162.
DR PROSITE; PS01095; CHITINASE_18.
KW GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 21
FT CHAIN 22 383 CARTILAGE GLYCOPROTEIN-39.
FT CARBOHYD 60 60 POTENTIAL.
SQ SEQUENCE 383 AA; 42613 MW; D0A5280F CRC32;

Query Match 44.0%; Score 1513; DB 4; Length 383;
Best Local Similarity 52.5%; Pred. No. 0.00e+00;
Matches 200; Conservative 89; Mismatches 83; Indels 9; Gaps 5;

```
Db      8 tgfvvlvllqccsayklvcyytswsqyregdgscfpdaldrflcthiiysfanisndhid 67
          :||:||:::   || |||||:|:|:|||:|:: :|  ||  ||||:||:|| ::| ::
Qy      8 AGFMVLLMIPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLS 67
```

```
Db      68 twewndvtlygmlntlknrnpnlktllsvggwnfgsqrfskiasntqsrrtfiksvppfl 127
Qy      68 TTEWNDETLYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFL 127

Db     128 rthgfdgldlawlypgrr-----dkqhfttlikemkaefikeaq-pgkkqlllsaalsag 181
Qy     128 RKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG 187

Db     182 kvtidssydiakisqhldfisimtydfhgawrgttghhsplfrgqedaspdrfsntdyav 241
Qy     188 QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAV 247

Db     242 gymlrlgapasklvmgiptfgrsftlass-etgvgapisgpgipgrftkeagtlayyeic 300
Qy     248 QQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVC 307

Db     301 dflrgatvhrtlgqqvpyatkgnqwvgyddqesvkskvqylkdrqlagamvwaldlddfq 360
Qy     308 SW-KGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFA 366

Db     361 gsfcgqdlrfpltnaikdala 381
Qy     367 GFSCNQG-RYPLIQTLRQELS 386
```

RESULT 2
ID CHIT_BRUMA STANDARD; PRT; 504 AA.
AC P29030;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE ENDOCHITINASE PRECURSOR (EC 3.2.1.14) (MF1 ANTIGEN).
OS BRUGIA MALAYI.
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; SPIRURIDA.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 92179220.
RA FUHRMAN J.A., LANE W.S., SMITH R.F., PIESSENS W.F., PERLER F.B.;
RL PROC. NATL. ACAD. SCI. U.S.A. 89:1548-1552(1992).
CC -!- FUNCTION: THE MF1 ANTIGEN IS A MICROFILARIAL CHITINASE, WHICH MAY
CC FUNCTION TO DEGRADE CHITIN-CONTAINING STRUCTURES IN THE MICRO-
CC FILARIA OR IN ITS MOSQUITO VECTOR DURING PARASITE DEVELOPMENT AND
CC TRANSMISSION.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC -!- PTM: O-GLYCOSYLATED.
CC -!- KNOWN TO BIND CALCIUM.
CC -!- DEVELOPMENTAL STAGE: THE APPEARANCE OF THE MF1 ANTIGEN CORRESPOND
CC WITH THE ONSET OF THE PARASITE'S ABILITY TO INFECT THE MOSQUITO.
CC -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; M73689; G156064; -.
DR PIR; A38221; A38221.
DR PROSITE; PS01095; CHITINASE_18.
KW HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ANTIGEN;
KW REPEAT; GLYCOPROTEIN; CALCIUM-BINDING.
FT SIGNAL 1 22
FT CHAIN 23 504 ENDOCHITINASE.
FT DOMAIN 23 400 CATALYTIC.
FT DOMAIN 401 450 SER/THR-RICH (LINKER).
FT DOMAIN 407 448 3 X 14 AA APPROXIMATE TANDEM REPEATS.
FT ACT_SITE 148 148 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 504 AA; 55971 MW; 4DA7E5EB CRC32;

Query Match 31.4%; Score 1080; DB 2; Length 504;
Best Local Similarity 42.5%; Pred. No. 4.93e-225;
Matches 157; Conservative 95; Mismatches 102; Indels 15; Gaps 11;

```
Db      27 cyytnwaqyrdgegkflpgnipnglcthilyafakvdelgdskpfewndedtewskgmys 86
Qy      26 CYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTT-EWNDE-T-----LYQ 78

Db      87 avtklretnpglkvllsyggynfgsaiftgiaksaqkterfiksaiaflrknnfdgfdld 146
Qy      79 EFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 138

Db     147 weyp-v-gv-a-e-eh-aklveamktafveeaktsgkqrllltaavsagkgtidgsynve 200
Qy     139 WEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVD 198

Db     201 slgknfdllflmsydlhgsweknvdlhgklhptkgevsgigifntefaadywaskgmpke 260
Qy     199 KIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPAS 258

Db     261 kiiigipmyaqgwtldnpsetaigaaasrpssasktnpaggtasyweickylkeggketv 320
Qy     259 KLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRI 317

Db     321 hqegvgaymvkgdqwygydneetirikmkwlkekgyggafiwaldfddftgkscgkgpyp 380
Qy     318 QDQKV-PYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYP 376

Db     381 llnaissel 389
Qy     377 LIQTLRQEL 385
```

RESULT 3
ID CHIT_MANSE STANDARD; PRT; 554 AA.
AC P36362;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
OS MANDUCA SEXTA (TOBACCO HAWKMOTH) (TOBACCO HORNWORM).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; LEPIDOPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93357793.
RA KRAMER K.J., CORPUZ L., CHOI H.K., MUTHUKRISHNAN S.;
RL INSECT BIOCHEM. MOL. BIOL. 23:691-701(1993).
CC -!- FUNCTION: DIGEST CHITIN IN THE EXOSKELETON DURING THE MOLTING
CC PROCESS.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC -!- DEVELOPMENTAL STAGE: HIGH LEVELS SEEN IN THE EPIDERMIS ON DAY 0,
CC BUT RAPIDLY DISAPPEARS AND IS UNDETECTED ON DAYS 1-4 OF FIFTH
CC INSTAR. IT REAPPEARS ON DAY 5 AND PEAKS ON DAY 7 AFTER WHICH A
CC RAPID DECLINE IS SEEN. IN THE GUT IS DETECTED ON DAY 6 WITH LOWER
CC LEVELS SEEN ON DAYS 0, 7 AND 8.
CC -!- TISSUE SPECIFICITY: EPIDERMIS AND GUT.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; U02270; G406049; -.
DR EMBL; S64757; -; NOT_ANNOTATED_CDS.
DR HSSP; P07254; 1CTN.
DR PROSITE; PS01095; CHITINASE_18.
KW HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 554 ENDOCHITINASE.
FT DOMAIN 396 453 SER/THR-RICH.
FT ACT_SITE 146 146 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 85 85 POTENTIAL.
FT CARBOHYD 303 303 POTENTIAL.
FT CARBOHYD 407 407 POTENTIAL.
FT CARBOHYD 545 545 POTENTIAL.
SQ SEQUENCE 554 AA; 62203 MW; FA87F8AD CRC32;

Query Match 29.8%; Score 1025; DB 2; Length 554;
Best Local Similarity 38.5%; Pred. No. 1.84e-211;
Matches 161; Conservative 106; Mismatches 126; Indels 25; Gaps 19;

```
Db      17 vqsdsrarivcyfsnwavyrpgvgrygiedipvekcthiiysfigvtegnsevliidpel 76
```

```
Qy      16 IPWGSAAKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTN-H-Q-LST-TEW 71

Db      77 dvdkngfrnftslrsshpsvkfmvavggwaegsskyshmvaqkstrmsfirsvvsflkky 136
           : |   :::| :|:  :| :| ::|:|||  :: |:: |||  : | :|: | : ||:||
Qy      72 N-DETLYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKY 130

Db     137 dfdgldldweypgaadrggsfsdkdkflylvqelrraf---irv-gkgwel-taavplan 191
           ||||||||||||:  :|::  ||::|  |||:|  ||    :  || : | :|||| :
Qy     131 SFDGLDLDWEYPGS--QGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQ 188

Db     192 frlmegyhvpelcqeldaihvmsydlrgnwagfadvhsplykrphdqwayeklnvndglh 251
             :  || |  : |:|| :::|:||::|:|   :  :|||||  :  |   |||: :::
Qy     189 TYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQ 248

Db     252 lweekgcpsnklvvgipfygrsftlsagnnnyglgtfinkeagggdpapytnatgfwayy 311
           | :|| |::||::|:| |||||||:::::   :|:     :|:| |:|:|:  |  |||
Qy     249 QWLQKGTPASKLILGMPTYGRSFTLASSSDT-RVGA--P-ATGSGTPGPFTKEGGMLAYY 304

Db     312 eictevdkddsgwtkkwdeqgkcpyaykgtqwvgyedprsveikmnwikqkgylgamtwa 371
           |:|:   |  :  || : :: | || ::  ||||::|  |   |:: :||||  ||| ||
Qy     305 EVCS--WKG-A--TKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWA 359

Db     372 idmddfqgl-cge-knplikilhkhms-sytvppphtenttptpewarppstpsdpse 426
           :|:||| |: | : : |||  |:  :| :| :|:   |   | |  :  |   : |::
Qy     360 LDLDDFAGFSCNQGRYPLIQTLRQELSLPY-LPSGTPELEVPKPGQPSEPEHGPSPGQ 416


RESULT    4
ID   CHI1_BACCI     STANDARD;      PRT;   699 AA.
AC   P20533;
DT   01-FEB-1991 (REL. 17, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE A1 PRECURSOR (EC 3.2.1.14).
GN   CHIA1.
OS   BACILLUS CIRCULANS.
OC   PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WL-12;
RX   MEDLINE; 90368776.
RA   WATANABE T., SUZUKI K., OYANAGI W., OHNISHI K., TANAKA H.;
RL   J. BIOL. CHEM. 265:15659-15665(1990).
RN   [2]
RP   MUTAGENESIS.
RC   STRAIN=WL-12;
RX   MEDLINE; 93366760.
RA   WATANABE T., KOHORI K., MIYASHITA K., FUJII T., SAKAI H.,
RA   UCHIDA M., TANAKA H.;
RL   J. BIOL. CHEM. 268:18567-18572(1993).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; M57601; G142688; -.
DR   PIR; A38368; A38368.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; REPEAT.
FT   SIGNAL        1     41
FT   CHAIN        42    699       CHITINASE A1.
FT   DOMAIN       42    460       CATALYTIC.
FT   DOMAIN      465    549       FIBRONECTIN TYPE-III (R-1).
FT   DOMAIN      560    644       FIBRONECTIN TYPE-III (R-2).
FT   ACT_SITE    204    204       PROTON DONOR (PROBABLE).
FT   MUTAGEN     200    200       D->N: DECREASE IN ACTIVITY.
FT   MUTAGEN     200    200       D->E: NO CHANGE IN ACTIVITY.
FT   MUTAGEN     204    204       E->D,Q: LOSS OF ACTIVITY.
SQ   SEQUENCE   699 AA;  73677 MW;  5A1777CC CRC32;

  Query Match           20.3%;  Score 700;  DB 2;  Length 699;
  Best Local Similarity 40.4%;  Pred. No. 3.64e-132;
  Matches  120;  Conservative   72;  Mismatches  85;  Indels  20;  Gaps  18;

Db     144 qlnklkqtnpnlktiisvggwtwsn-rfsdvaataatrevfansavdflrkynfdgvdld 202
           ::| ||  ||:|||:::::|||  |  :|:|: |||  |: | |||: |||||:|||:|||
Qy      79 EFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 138

Db     203 weypvsggldgnskrpedkqnytlllskirekldaagavdgk-kylltiasgasatya-a 260
           ||||  |: :| |  : ||: :| |:  : : ::  :   || : ||: |  |: ||  |
Qy     139 WEYP--GS-QG-S-PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDA 193

Db     261 ntelakiaaivdwinimtydfngawqkisahnaplnydpaasaagvpdantfnvaagaqg 320
             |: |||  :| :|:|:|||:|:|:|:::||:|| |      :| : | ::|| |: |
Qy     194 GYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPL-YKRQE-ESGAA-A-SLNVDAAVQQ 249

Db     321 hldagvpaaklvlgvpfygrgwd-gcaqagn-gqyqtctggssvgtweagsfdfydlean 378
           |: | ||:||:||:| |||:   : :     |   | :| ::  | |:| : :|::  :
Qy     250 WLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEV-CS 308

Db     379 yinkngytrywndtakvpylynasnkrfisyddaesvgyktayikskglggamfwel 435
             | | |: : :  ||||::   | : :::|| ||   | :|:| |||||||  | |
Qy     309 W--K-GATKQRIQDQKVPYIFRD-NQW-VGFDDVESFKTKVSYLKQKGLGGAMVWAL 360


RESULT    5
ID   CHI1_APHAL     STANDARD;      PRT;   423 AA.
AC   P32470;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   CHITINASE 1 PRECURSOR (EC 3.2.1.14).
GN   CHI1.
OS   APHANOCLADIUM ALBUM.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ETHM 483;
RX   MEDLINE; 93013040.
RA   BLAISEAU P.-L., LAFAY J.-F.;
RL   GENE 120:243-248(1992).
RN   [2]
RP   SEQUENCE OF 35-57.
RX   MEDLINE; 92136437.
RA   BLAISEAU P.-L., KUNZ C., GRISON R., BERTHEAU Y., BRYGOO Y.;
RL   CURR. GENET. 21:61-66(1992).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; X64104; G429026; -.
DR   PIR; JQ1975; JQ1975.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ZYMOGEN.
FT   SIGNAL        1     22       POTENTIAL.
FT   PROPEP       23     34       POTENTIAL.
FT   CHAIN        35    423       CHITINASE 1.
FT   ACT_SITE    171    171       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   423 AA;  46058 MW;  88456E6E CRC32;

  Query Match           16.8%;  Score 579;  DB 2;  Length 423;
  Best Local Similarity 36.6%;  Pred. No. 3.08e-103;
  Matches  102;  Conservative   64;  Mismatches  95;  Indels  18;  Gaps  16;

Db     115 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 173
           ||| | ::| :|:|||| || :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy      83 LKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 142

Db     174 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 228
            :::: | ::  :  |:  ::: :  : |  ::||| | ||| : |   :  : ::
Qy     143 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 201

Db     229 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 287
```

```
           ||::|||||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy     202 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 261

Db     288 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 339
           |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy     262 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 321

Db     340 gcysydpstkelisfdtpamistkvswlkgkglggtmfw 378
             | :    :: ::||    : |||| || |||||:| |
Qy     322 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT    6
ID   CHI4_TRIHA     STANDARD;      PRT;   423 AA.
AC   P48827;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   42 KD ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
GN   CHIT42.
OS   TRICHODERMA HARZIANUM.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 35-52; 93-107; 371-385 & 397-414.
RX   MEDLINE; 95269313.
RA   GARCIA I., LORA J.M., LA CRUZ J., BENITEZ T., LLOBELL A.,
RA   PINTOR-TORO J.A.;
RL   CURR. GENET. 27:83-89(1994).
CC   -!- FUNCTION: MORPHOGENETIC ROLE DURING APICAL GROWTH,CELL DIVISION
CC       AND DIFFERENTIATION (CELL WALL MORPHOGENESIS). ANTIFUNGAL
CC       AGENT.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- INDUCTION: SPECIFICALLY INDUCED BY CHITIN AND IS CATABOLITE
CC       REPRESSED.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; S78423; G999376; -.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ZYMOGEN;
KW   CHITIN-BINDING.
FT   SIGNAL        1     22       POTENTIAL.
FT   PROPEP       23     34
FT   CHAIN        35    423       42 KD ENDOCHITINASE.
FT   ACT_SITE    171    171       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    218    218       POTENTIAL.
SQ   SEQUENCE   423 AA;  46056 MW;  899DA50A CRC32;

  Query Match           16.4%;  Score 564;  DB 2;  Length 423;
  Best Local Similarity  36.0%;  Pred. No. 1.10e-99;
  Matches   102;  Conservative   70;  Mismatches  94;  Indels  17;  Gaps  13;

Db     111 qlfk-vkkanrglkvllsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidi 168
           | |: :|| |  || ||:|||  || :|    :|  ||  | ::|| |::  :|||:|:
Qy      78 QEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 137

Db     169 dweyp-adatqasnmilllkevrsqrdayaaqy-apg-yhflltiaapagkdny-sklrl 224
           |||||  :::: | ::  :   |:  :|:  :  ::|  ::||: | |||    :   :
Qy     138 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 197

Db     225 adlgqvldyinlmaydyagsfspltghdanlfnnpsnpnatp-fntdsavkdyinggvpa 283
             ::| ||::|||||: ||   :|||:: |::   :: |:: :| |:|| : :  | ||
Qy     198 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 257

Db     284 nkivlgmpiygrsf--qntag--igqtyngvgs-gsw--eagiwdykalpk-agatvqyd 335
           :|::|||| |||||   :::   :|    | |: |:   |:|:  |  :    ||| |
Qy     258 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRI 317

Db     336 svakgyysynsatkelisfdtpdmintkvaylkslglggsmfw 378
             |  | :    :: ::||  : ::|||:|||  ||||:| |
Qy     318 QDQKVPYIFRD--NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT    7
ID   CHI1_COCIM     STANDARD;      PRT;   427 AA.
AC   P54196;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   ENDOCHITINASE 1 PRECURSOR (EC 3.2.1.14) (COMPLEMENT-FIXATION ANTIGEN)
DE   (CF-ANTIGEN) (CF-AG).
GN   CTS1.
OS   COCCIDIOIDES IMMITIS.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C735;
RX   MEDLINE; 96144270.
RA   PISHKO E.J., KIRKLAND T.N., COLE G.T.;
RL   GENE 167:173-177(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SILVEIRA;
RA   YANG C., ZHU Y., MAGEE D.M., COX R.A.;
RL   SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; L41663; G1200190; -.
DR   EMBL; U51271; G1256769; -.
DR   EMBL; U33265; G1255728; -.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL;
KW   GLYCOPROTEIN.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?    427       ENDOCHITINASE 1.
FT   CARBOHYD    387    387       POTENTIAL.
FT   CONFLICT     15     47       RWLSLRCLFCELGRFMFTLSTVTAVTVVTDDIQ -> VQAS
FT                                SMSSMPNYYPVPEAPEGGFRSVVYFVNW (IN REF. 2).
FT   CONFLICT    199    199       K -> N (IN REF. 2).
SQ   SEQUENCE   427 AA;  47629 MW;  6CB9AF73 CRC32;

  Query Match           15.9%;  Score 548;  DB 2;  Length 427;
  Best Local Similarity  34.4%;  Pred. No. 6.62e-96;
  Matches    96;  Conservative   70;  Mismatches  94;  Indels  19;  Gaps  17;

Db     115 lkknnrnlktllsiggwtysp-nfktpasteegrkkfadtslklmkdlgfdgididweyp 173
           ||| | :|||||:|||| :|  :|    :| : |  | :::::::  :|||:|:|||||
Qy      83 LKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 142

Db     174 ede-kqandfvlllkacrealdaysak-hpngk-kflltiaspagpqny-nk-lklaemd 228
           :   | |   :    ::  :|:    : :|| ::||: | ||| | | :    :  :
Qy     143 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 201

Db     229 kyldfwnlmaydfsgswdkvsghmsnvfpsttkp-estpfssdkavkdyikagvpankiv 287
           ||| ||||||| |||:||:|| | ::      :  ::::: | || : :  | ||:|::
Qy     202 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 261

Db     288 lgmplygraf--ast-dg-igtsfngvgg-gsw--engvwdykdmpq-qgaqvteledia 339
           |||| |||:|  ||: |  :|::  | |  |:   | |:  | ::    ||    ::|
Qy     262 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 321

Db     340 asysydknkrylisydtvkiagkkaeyitkngmgggm-w 377
           :| : ::::  :::| |      |  |:  :|:||:| |
Qy     322 VPYIF-RDNQW-VGFDDVESFKTKVSYLKQKGLGGAMVW 358


RESULT    8
ID   CHIA_SERMA     STANDARD;      PRT;   563 AA.
AC   P07254;
DT   01-APR-1988 (REL. 07, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE A PRECURSOR (EC 3.2.1.14).
GN   CHIA.
OS   SERRATIA MARCESCENS.
```

```
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   ENTEROBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KOO J.C., LIM C.O., CHOI Y.J., KIM C.Y., BAHK J.D., LEE S.Y.,
RA   CHO M.J.;
RL   SUBMITTED (JAN-1993) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 990 / QMB1466;
RA   JONES J.D.G., GRADY K.L., SUSLOW T.V., BEDBROOK J.R.;
RL   EMBO J. 5:467-473(1986).
RN   [3]
RP   REVISIONS, AND X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE; 95219379.
RA   PERRAKIS A., TEWS I., DAUTER Z., OPPENHEIM A.B., CHET I., WILSON K.S.,
RA   VORGIAS C.E.;
RL   STRUCTURE 2:1169-1180(1994).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; L01455; G152818; -.
DR   EMBL; X03657; G46831; -.
DR   PIR; A25090; A25090.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL.
FT   SIGNAL        1     23
FT   CHAIN        24    563       CHITINASE A.
FT   DOMAIN      150    563       CATALYTIC.
FT   ACT_SITE    315    315       PROTON DONOR (PROBABLE).
FT   ACT_SITE    391    391       PROBABLE.
FT   CONFLICT     76     77       TA -> GP (IN REF. 2).
FT   CONFLICT    139    139       A -> P (IN REF. 2).
FT   CONFLICT    226    226       V -> I (IN REF. 2).
FT   CONFLICT    395    395       P -> A (IN REF. 1 AND 2).
FT   CONFLICT    410    430       PAWKPDTAYTTVNGVNALLAQ -> RPGSRHRLHHGERRQC
FT                                AAGQ (IN REF.1 AND 2).
FT   CONFLICT    437    437       V -> I (IN REF. 2).
FT   CONFLICT    464    467       ATGP -> HRA (IN REF. 2).
FT   CONFLICT    473    473       K -> E (IN REF. 2).
FT   CONFLICT    484    484       G -> S (IN REF. 2).
SQ   SEQUENCE   563 AA;  60979 MW;  7764DB57 CRC32;

  Query Match           12.8%;  Score 440;  DB 2;  Length 563;
  Best Local Similarity  29.1%;  Pred. No. 1.04e-70;
  Matches    94;  Conservative   83;  Mismatches 123;  Indels  23;  Gaps  19;

Db    231 pfaalqkaqkgvtawddpykgnfgqlmalkqahpdlkilpsiggwtlsdpfff-mgdkvk 289
          :||:: : | : | :|     : :: :||  :| || | :|||| :|   | |    :
Qy     57 AFAGMTNHQLSTTEWNDE-T-LYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATAN 114

Db    290 -rdrfvgsvkeflqtwkffdgvdidwefpggkganpnlgspqdgetyvllmkelramldq 348
           |: || |   ||:   | ||:|:|||:||: |: | :    | | :  |: :|   ::|
Qy    115 NRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV----DKERFTTLVQDLANAFQQ 168

Db    349 lsaetgr-kyeltsaisagkdkidkvaynvaq--nsmdhiflmsydfygpfdlknlghqt 405
          :  :|: :  |::|::||   :|  :|:|     ::| : ||:||| |: : |  || :
Qy    169 EAQTSGKERLLLSAAVPAGQTYVD-AGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNS 226

Db    406 alnapawkpdtayt-tvng-vnallaqgvkpgkvvvgtamygrgwtgvngyqnnipftgt 463
          :|      : :| :  |:: |   |  |  ::|:::| : |||: |  :: :  :  :|
Qy    227 PLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPAT 286

Db    464 atg-pvkgtwkngivdyrqiagqfmsgewqytydataeapyvfkpstgdlitfddarsvq 522
          ::| |   |   |:: | :: :  :::  |   :     ||:|:    : : |||  |
Qy    287 GSGTPGPFTKEGGMLAYYEVCS-WKGATKQ-RIQD-QKVPYIFR-DN-QWVGFDDVESFK 341

Db    523 akgkyvldkqlgglfsweidadn 545
          :|  |: :| |||   | :| |:
Qy    342 TKVSYLKQKGLGGAMVWALDLDD 364
```

```
RESULT    9
ID   CHIT_NPVAC     STANDARD;      PRT;   551 AA.
AC   P41684;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   PROBABLE ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
OS   AUTOGRAPHA CALIFORNICA NUCLEAR POLYHEDROSIS VIRUS (ACMNPV).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; BACULOVIRIDAE; EUBACULOVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C6;
RX   MEDLINE; 94303173.
RA   AYRES M.D., HOWARD S.C., KUZIO J., LOPEZ-FERBER M., POSSEE R.D.;
RL   VIROLOGY 202:586-605(1994).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; L22858; G559195; -.
DR   PROSITE; PS00014; ER_TARGET.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    551       PROBABLE ENDOCHITINASE.
FT   ACT_SITE    305    305       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    173    173       POTENTIAL.
FT   CARBOHYD    444    444       POTENTIAL.
FT   SITE        548    551       PREVENT SECRETION FROM ER (POTENTIAL).
SQ   SEQUENCE   551 AA;  61368 MW;  BDB124DB CRC32;

  Query Match           11.6%;  Score 398;  DB 2;  Length 551;
  Best Local Similarity  29.5%;  Pred. No. 4.28e-61;
  Matches    96;  Conservative   79;  Mismatches 122;  Indels  28;  Gaps  27;

Db    221 pwaavqkpqkgvsawn-ep-ykgnfgqlmaaklanphlkilpsiggwtlsdp-fyfmhdv 277
          : |:: : | : : || |  |  :|  |:  |: || || | :|||| :|   |  |
Qy     57 AFAGMTNHQLSTTEWNDETLYQ-EFNGLK--KM-NPKLKTLLAIGGWNFSTQKFTDMVAT 112

Db    278 -ekrnvfvdsvkeflqvwkffdgvdidwefpggkganpslgdadgdaktyillleelram 336
           ::|  ||:|   ||:   | ||:|:|||:||: |: |:: |   : : :  |:::|
Qy    113 ANNRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV-D-K-E-R-FTTLVQDLANA 165

Db    337 lddleaqt-grv-yeltsaisagydkiavvnya-e-aqkslgkiflmsydfkgawsntdl 392
          ::: |||| |:    |::|::|| :       |  :    :|  : ||:||| |:| :
Qy    166 FQQ-EAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVT- 222

Db    393 gyqttvya-ps-wnseelytthyavdallkqgvdpnkiivgvamygrgwtgvtnytndny 450
          |  : :|       :          ||:  |  |  ::|:|:|:: |||: |  ::
Qy    223 GHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVG 282

Db    451 fsgtgng-pgsgtwedgvvdyrqiqkdlnnyvytfdsaaqasyvfdkskgdlisfdsvds 509
          ::||:| ||: | | |:: | ::          ::     :|:| : : : ::|| |:|
Qy    283 APATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQ-KVPYIF-RDN-QWVGFDDVES 339

Db    510 vlgkvkyvdrnklgglfaweidadn 534
             || |: :: |||   | :| |:
Qy    340 FKTKVSYLKQKGLGGAMVWALDLDD 364


RESULT   10
ID   CHIA_ALTSO     STANDARD;      PRT;   820 AA.
AC   P32823;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE A PRECURSOR (EC 3.2.1.14) (CHI-A).
GN   CHIA.
OS   ALTEROMONAS SP. (STRAIN O-7).
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   VIBRONACEAE.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93106952.
RA   TSUJIBO H., ORIKOSHI H., TANNO H., FUJIMOTO K., MIYAMOTO K.,
RA   IMADA C., OKAMI Y., INAMORI Y.;
RL   J. BACTERIOL. 175:176-181(1993).
RN   [2]
RP   SEQUENCE OF 22-48, AND CHARACTERIZATION.
RX   MEDLINE; 93099552.
RA   TSUJIBO H., YOSHIDA Y., MIYAMOTO K., IMADA C., OKAMI Y., INAMORI Y.;
RL   CAN. J. MICROBIOL. 38:891-897(1992).
RN   [3]
RP   MUTAGENESIS.
RX   MEDLINE; 93379363.
RA   TSUJIBO H., ORIKOSHI H., IMADA C., OKAMI Y., MIYAMOTO K.,
RA   INAMORI Y.;
RL   BIOSCI. BIOTECHNOL. BIOCHEM. 57:1396-1397(1993).
CC   -!- THE OPTIMUM PH AND TEMPERATURE OF CHI-A ARE 8.0 AND 50 DEGREES
CC       CELSIUS, RESPECTIVELY.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- ENZYME REGULATION: STIMULATED BY MG2+; INHIBITED BY
CC       N-BROMOSUCCINIMIDE AND 2-HYDROXY-5-NITROBENZYL BROMIDE.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; D13762; G216207; -.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL.
FT   SIGNAL        1     21
FT   CHAIN        22    820       CHITINASE A.
FT   ACT_SITE    313    313       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   820 AA;  87346 MW;  838F6EE8 CRC32;

  Query Match            10.1%;  Score 348;  DB 2;  Length 820;
  Best Local Similarity  35.6%;  Pred. No. 7.81e-50;
  Matches    58;  Conservative   43;  Mismatches  47;  Indels  15;  Gaps  13;

Db    250 vysqlmalkqrypdlkilpsvggwtlsdp-fhgftnkan-rdtfvasvkqflktwkfydg 307
          :| :: :||   | || | ::||| :|   |      || |:||| |  :||:   | ||
Qy     76 LYQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYSF-DG 134

Db    308 vdidwefpggdgpnpdlgdpindgpayvalmqelramldeleaet-grq-yeltsaigag 365
          :|:|||:||::|: | : |  ::   : :|:|:|   ::: ||:| |::   |::|: ||
Qy    135 LDLDWEYPGSQGS-PAV-D--KE-R-FTTLVQDLANAFQQ-EAQTSGKERLLLSAAVPAG 187

Db    366 ydkiedvdyqaaq--qymdyifamtydfygawnnetghqtgiy 406
          :   |  |:     | :|::  |:||| |:|:: ||| : :|
Qy    188 -QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLY 229


RESULT   11
ID   CHIB_SERMA     STANDARD;      PRT;   499 AA.
AC   P11797;
DT   01-OCT-1989 (REL. 12, CREATED)
DT   01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE B PRECURSOR (EC 3.2.1.14).
GN   CHIB.
OS   SERRATIA MARCESCENS.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   ENTEROBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 990 / QMB1466;
RX   MEDLINE; 89345110.
RA   HARPSTER M.H., DUNSMUIR P.;
RL   NUCLEIC ACIDS RES. 17:5395-5395(1989).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; X15208; G47228; -.
DR   PIR; S04856; S04856.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL.
FT   SIGNAL        1     41
FT   CHAIN        42    499       CHITINASE B.
FT   ACT_SITE    144    144       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   499 AA;  55464 MW;  9A585FFC CRC32;

  Query Match             9.7%;  Score 333;  DB 2;  Length 499;
  Best Local Similarity  33.3%;  Pred. No. 1.68e-46;
  Matches    53;  Conservative   47;  Mismatches  45;  Indels  14;  Gaps   8;

Db     78 ltalkahnpslrimfsiggwyysndlgvshanyvnavktpaartkfaqscvrimkdygfd 137
          : :||  || |: :::|||| :|     :: :: : | |:  |  |  | :|::: |:||
Qy     80 FNGLKKMNPKLKTLLAIGGWNFS-----TQ-KFTDMVATANNRQTFVNSAIRFLRKYSFD 133

Db    138 gvdidweyp--qaa-evd-g-fiaalqeirtllnqqtiadgrqalpyqltiagaggaffl 192
          |:|:|||||  |::  ||   | : :|::   : |:: : |:: |   |: | ::|  ::
Qy    134 GLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERL--LLSAAVPAGQTYV 191

Db    193 sryysklaqivapldyinlmtydlagpwekitnhqaalf 231
          |  :  |   ||::|||:||: |:|||:| | ::|:
Qy    192 DAGYE-VDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLY 229


RESULT   12
ID   CHIT_STRLI     STANDARD;      PRT;   619 AA.
AC   P36909;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE C PRECURSOR (EC 3.2.1.14).
GN   CHIC.
OS   STREPTOMYCES LIVIDANS.
OC   PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93294525.
RA   FUJII T., MIYASHITA K.;
RL   J. GEN. MICROBIOL. 139:677-686(1993).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- INDUCTION: BY CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS A BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC       (CBD).
DR   EMBL; D12647; G391910; -.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS00018; EF_HAND.
DR   PROSITE; PS00561; CBD_BACTERIAL.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    619       CHITINASE C.
FT   DOMAIN       35    140       CELLULOSE-BINDING.
FT   DOMAIN      148    230       FIBRONECTIN TYPE-III.
FT   DOMAIN      240    619       CATALYTIC.
FT   ACT_SITE    382    382       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   619 AA;  65200 MW;  23E3E1EC CRC32;

  Query Match             8.8%;  Score 304;  DB 2;  Length 619;
  Best Local Similarity  28.5%;  Pred. No. 3.91e-40;
  Matches    87;  Conservative   64;  Mismatches 129;  Indels  25;  Gaps  18;

Db    311 w-dqplrgnfnqlrklkakypnikilysfggwtwsgg-fpdavknpaafakschdlvedp 368
          | |: |  :|| |:|::   |::| | ::|||  |   | | |  :       :  :
Qy     71 WNDETLYQEFNGLKKMN---PKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFL 127

Db    369 rwadvfdgidldweypnacglscdetsapnafssmmkamraefgqdylita-a---vtad 424
          |    |||:|||||| : | |   :   : |:::  :   | |:    ::      ::|
Qy    128 RKYS-FDGLDLDWEYPGSQG-S--PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAA 183
```

```
Db    425 gsdggk-idaa-dygeaskyidwynvmtydffgawak-ngptaphspltaydgipqqgfn 481
          : |   :||: :    : :|  |:|:||| |:| |  |  :|       | :  ::|
Qy    184 VPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA-SLN 242

Db    482 tadamakfkskgvpadklligigfygrgwt-gvtqsap-ggtatgpat-gty--eagied 536
            |:     || || ||::|:  |||: | : :  :  |: |||::| | :  |:|:
Qy    243 VDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLA 302

Db    537 y-kvlknscpatgtiag--tayahcgsnwwsydtpatikskmdwaeqqglggaffwefsg 593
          | |     ::   |     :|    : | ::|   ::|:|:    | |||||  | :
Qy    303 YYEVCSWKGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDL 362

Db    594 dtang 598
          | |
Qy    363 DDFAG 367


RESULT   13
ID   CHIT_STRPL     STANDARD;      PRT;   610 AA.
AC   P11220;
DT   01-JUL-1989 (REL. 11, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE 63 PRECURSOR (EC 3.2.1.14).
GN   CHTA.
OS   STREPTOMYCES PLICATUS.
OC   PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92192480.
RA   ROBBINS P.W., OVERBYE K., ALBRIGHT C., BENFIELD B., PERO J.;
RL   GENE 111:69-76(1992).
RN   [2]
RP   SEQUENCE OF 1-45 FROM N.A., AND SEQUENCE OF 31-45.
RX   MEDLINE; 88087127.
RA   ROBBINS P.W., ALBRIGHT C., BENFIELD B.;
RL   J. BIOL. CHEM. 263:443-447(1988).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- INDUCTION: BY CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS A BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC       (CBD).
DR   EMBL; M82804; G153216; -.
DR   EMBL; M18397; G153209; -.
DR   PIR; A29912; A29912.
DR   PIR; JH0573; JH0573.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS00018; EF_HAND.
DR   PROSITE; PS00561; CBD_BACTERIAL.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL.
FT   SIGNAL        1     30
FT   CHAIN        31    610       CHITINASE 63.
FT   DOMAIN       35    140       CELLULOSE-BINDING.
FT   DOMAIN      148    229       FIBRONECTIN TYPE-III.
FT   DOMAIN      236    610       CATALYTIC.
FT   ACT_SITE    383    383       PROTON DONOR (BY SIMILARITY).
FT   CONFLICT      3      3       F -> I (IN REF. 2).
SQ   SEQUENCE   610 AA;  63974 MW;  2F5E8E35 CRC32;

  Query Match             8.4%;  Score 290;  DB 2;  Length 610;
  Best Local Similarity  26.7%;  Pred. No. 4.24e-37;
  Matches    79;  Conservative   62;  Mismatches 134;  Indels  21;  Gaps  16;

Db    320 fnqlrnlkaeyphikilysfggwtwsgg-fpdavknpaafakschdlvedprwadvfdgi 378
          : ::  ||   | :| | ::|||  |   | | |  :       :  :   |    |||:
Qy     77 YQEFNGLKKMNPKLKTLLAIGGWNFSTQKFTDMVATANNRQTFVNSAIRFLRKYS-FDGL 135

Db    379 dldweypnacglscdetsapnafssmmkamraefgqdylita-a---vtadgsdggk-id 433
          |||||| : | |   :   : |:::: :   | |:   ::     ::|  : |   :|
Qy    136 DLDWEYPGSQG-S--PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVD 192

Db    434 aa-dygeaskyidwynvmtydffgawak-ngptaphsplnaydgipqqgfttadamakfk 491
          |: :    :  :|  |:|:||| |:| |  |  :|       | :  ::    |:
Qy    193 AGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA-SLNVDAAVQQWL 251

Db    492 skgvpadklligigfygrgwt-gvtqsap-ggtatgpaa-gty--eagiedy-kvlknsc 545
            || || ||::|:  |||: | : :  :  |: |||::: | :  |:|:  |  |
Qy    252 QKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKG 311

Db    546 patgtvag--tayahcgtnwwsydtpatikskmdwaeqqglggaffwefsgdttng 599
          ::   :     :|      | ::|   ::|:|:    | |||||  | :  |   |
Qy    312 ATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAG 367


RESULT   14
ID   KTXA_KLULA     STANDARD;      PRT;  1146 AA.
AC   P09805;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   KILLER TOXIN ALPHA AND BETA SUBUNITS PRECURSOR (RF2 PROTEIN)
DE   (ENDOCHITINASE (EC 3.2.1.14)).
OS   KLUYVEROMYCES LACTIS (YEAST).
OG   PLASMID PGKL-1.
OC   EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS 2359/152;
RA   SOR F., FUKUHARA H.;
RL   CURR. GENET. 9:147-155(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84297209.
RA   STARK M.J.R., MILEHAM A.J., ROMANOS M.A., BOYD A.;
RL   NUCLEIC ACIDS RES. 12:6011-6030(1984).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85037931.
RA   TOKUNAGA M., KAWAMURA A., HISHINUMA F.;
RL   NUCLEIC ACIDS RES. 12:7581-7597(1984).
RN   [4]
RP   IDENTIFICATION OF PROTEIN, AND SEQUENCE OF 30-44 AND 895-916.
RX   MEDLINE; 87004569.
RA   STARK M.J.R., BOYD A.;
RL   EMBO J. 5:1995-2002(1986).
RN   [5]
RP   SIMILARITY TO CHITINASE OF ALPHA-SUBUNIT.
RX   MEDLINE; 90259069.
RA   BRADSHAW H.D. JR.;
RL   NATURE 345:299-299(1990).
RN   [6]
RP   CHITINASE ACTIVITY OF ALPHA-SUBUNIT.
RX   MEDLINE; 91301161.
RA   BUTLER A.R., O'DONNEL R.W., MARTIN V.J., GOODAY G.W., STARK M.J.R.;
RL   EUR. J. BIOCHEM. 199:483-488(1991).
CC   -!- FUNCTION: THE ALPHA SUBUNIT IS A POTENT EXOCHITINASE. ALONG WITH
CC       THE BETA SUBUNIT IT PLAYS A ROLE IN THE INITIAL INTERACTION OF THE
CC       TOXIN WITH SENSITIVE CELLS AND ALLOW THE GAMMA SUBUNIT (THE ACTIVE
CC       TOXIN) TO GAIN ENTRY INTO THE CELL.
CC   -!- PTM: RF2 IS POTENTIALLY SPLIT BY MEMBRANE-BOUND BASIC AMINO ACID-
CC       SPECIFIC PEPTIDASE TO YIELD THE ALPHA AND BETA SUBUNITS.
CC   -!- SUBUNIT: THE KILLER TOXIN IS COMPOSED OF THREE SUBUNITS: ALPHA,
CC       BETA AND GAMMA.
CC   -!- SIMILARITY: LOCAL, TO OTHER CHITIN-BINDING PROTEINS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
DR   EMBL; X07127; G2829; -.
DR   EMBL; X00762; G2844; -.
DR   EMBL; X01095; G2849; -.
DR   PIR; S07915; S07915.
DR   HSSP; P02877; 1HEV.
DR   PROSITE; PS00026; CHITIN_BINDING.
```

```
DR   PROSITE; PS01095; CHITINASE_18.
KW   PLASMID; TOXIN; GLYCOPROTEIN; SIGNAL; HYDROLASE; GLYCOSIDASE;
KW   CHITIN DEGRADATION.
FT   SIGNAL        1     17       POTENTIAL.
FT   PROPEP       18     29       POTENTIAL.
FT   CHAIN        30    892       ALPHA SUBUNIT (PROBABLE).
FT   CHAIN       895   1146       BETA SUBUNIT (PROBABLE).
FT   DOMAIN      310    363       CHITIN-BINDING (BY SIMILARITY).
FT   DOMAIN      364    635       CATALYTIC.
FT   ACT_SITE    495    495       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    771    771       POTENTIAL.
FT   CARBOHYD    858    858       POTENTIAL.
FT   CARBOHYD    868    868       POTENTIAL.
FT   CARBOHYD    876    876       POTENTIAL.
FT   CARBOHYD    995    995       POTENTIAL.
FT   CARBOHYD   1082   1082       POTENTIAL.
FT   CARBOHYD   1117   1117       POTENTIAL.
SQ   SEQUENCE   1146 AA;  128936 MW;  CDD1620F CRC32;

  Query Match             7.7%;  Score 266;  DB 5;  Length 1146;
  Best Local Similarity  28.9%;  Pred. No. 5.77e-32;
  Matches    61;  Conservative   57;  Mismatches  74;  Indels  19;  Gaps  14;

Db    425 ddsafsksaflkvtssk-k-ipsfggwdfstspstytifrnavktdqnrntfannlinfm 482
          |:: :     ||  ::| | : ::|||:|||    :|   : | |  || || |: | |:
Qy     73 DETLYQEFNGLKKMNPKLKTLLAIGGWNFST-Q-KFT---DMVATANNRQTFVNSAIRFL 127

Db    483 nkynldgidldweypgapdipdipaddsssgsnyltflkllkgkmps-gktl-siaipss 540
           ||::||:||||||||:   | :  :  ::    |:      :  ::  : | | |:|::
Qy    128 RKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG 187

Db    541 ywylkn-fpisdiqntvdymvymtydihgiwe--ygkans-yi---nchtprk-eiedai 592
            |:   : :  |   :|::  |:||:|| ||   |  :: |    :  ::   ::: |:
Qy    188 QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAV 247

Db    593 km-ldkagvkfnkvfggvanygrsykmvntn 622
             |:| |   :|:: |:: |||| : :::
Qy    248 QQWLQK-GTPASKLILGMPTYGRSFTLASSS 277


RESULT   15
ID   GP39_BOVIN     STANDARD;      PRT;    36 AA.
AC   P30922;
DT   01-JUL-1993 (REL. 26, CREATED)
DT   01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   CARTILAGE GLYCOPROTEIN-39 (GP-39) (39 KD WHEY PROTEIN) (FRAGMENT).
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; ARTIODACTYLA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 88106603.
RA   REJMAN J.J., HURLEY W.L.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 150:329-334(1988).
CC   -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE CAPACITY OF CELLS TO
CC       RESPOND TO AND COPE WITH CHANGES IN THEIR ENVIRONMENT.
CC   -!- PTM: GLYCOSYLATED.
CC   -!- TISSUE SPECIFICITY: MAMMARY SECRETIONS COLLECTED DURING THE
CC       NONLACTATING PERIOD.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
DR   PIR; A27682; A27682.
KW   GLYCOPROTEIN.
FT   NON_TER      36     36
SQ   SEQUENCE   36 AA;  4264 MW;  F3935740 CRC32;

  Query Match             5.0%;  Score 172;  DB 4;  Length 36;
  Best Local Similarity  48.6%;  Pred. No. 4.73e-13;
  Matches    17;  Conservative   11;  Mismatches   7;  Indels   0;  Gaps   0;

Db      2 klixyytswsqyregdgsxfpdaidpflxthviys 36
          ||: |:|:|:|||:|::  :|  :|| | ||:||:
Qy     23 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYA 57


Search completed: Fri Jun 27 16:55:57 1997
Job time : 96 secs.
```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 27 17:03:11 1997; MasPar time 9.60 Seconds
423.242 Million cell updates/sec

Tabular output not generated.

Title: >US-08-663-618A-14
Description: (1-373) from US08663618A.pep
Perfect Score: 2721
Sequence: 1 AKLVCYFTNWAQYRQGEARF.....YPLIQTLRQELSLPYLPSGT 373

Scoring table: PAM 150
Gap 11

Searched: 92623 seqs, 10896596 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq26
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19

Statistics: Mean 34.970; Variance 153.103; scale 0.228

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1328	48.8	537	14	R73991	Bovine oviduct specif	7.85e-114
2	1303	47.9	668	14	R73993	Hamster oviduct speci	2.08e-111
3	1303	47.9	718	14	R73992	Murine oviduct specif	2.08e-111
4	881	32.4	321	1	P81342	Polypeptide involved	9.90e-71
5	597	21.9	371	13	R70025	Tobacco chitinase enc	9.86e-44
6	597	21.9	377	13	R70029	Tobacco cluster-A pro	9.86e-44
7	581	21.4	424	12	R63441	Trichoderma harzianum	3.16e-42
8	576	21.2	389	10	R56861	Aphanocladium album m	9.31e-42
9	576	21.2	423	6	R33072	Aphanocladium album p	9.31e-42
10	569	20.9	389	6	R33068	Aphanocladium album m	4.23e-41
11	463	17.0	866	19	W02159	Soluble chitinase.	3.32e-31
12	389	14.3	561	12	R64823	Chitinase A.	2.18e-24
13	372	13.7	561	9	R48981	Bacterial chitinase-A	7.79e-23
14	271	10.0	238	6	R32547	HaNPV IE-1 gene prote	9.19e-14
15	219	8.0	1046	19	W02156	Periplasmic chitodext	3.07e-09
16	139	5.1	25	13	R70751	YKL-40 N-terminal pep	1.21e-02
17	139	5.1	25	12	R70745	YKL-40 N-terminal seq	1.21e-02
18	136	5.0	29	11	R60073	Bovine Oviduct Glycop	2.08e-02
19	101	3.7	1070	4	R21521	Alpha galactosidase.	8.46e+00
20	99	3.6	398	18	R95856	S. pyogenes speB gene	1.17e+01
21	99	3.6	428	16	R86905	Murine BMP-9.	1.17e+01
22	99	3.6	428	6	R31372	Murine bone morphogen	1.17e+01
23	99	3.6	553	14	R80445	Murine soluble epoxid	1.17e+01
24	99	3.6	958	1	P81181	Sequence of glucoamyl	1.17e+01
25	99	3.6	958	2	R07575	Glucoamylase encoded	1.17e+01
26	95	3.5	438	9	R47259	Pre-pro-VGR1.	2.22e+01
27	92	3.4	1356	5	R26999	Novel type III RTK en	3.57e+01
28	91	3.3	19	13	R70752	YKL-40 internal pepti	4.18e+01
29	91	3.3	19	12	R70746	YKL-40 internal pepti	4.18e+01
30	90	3.3	256	17	R97235	MuTu putative oncogen	4.88e+01
31	90	3.3	459	17	R88058	Protein encdoed by Mu	4.88e+01
32	90	3.3	575	3	P50018	Sequence encoded by t	4.88e+01
33	90	3.3	709	2	P70423	FixD nitrogen fixatio	4.88e+01
34	90	3.3	709	3	P50017	Sequence encoded by t	4.88e+01
35	87	3.2	369	18	R98125	Human lymphocyte cell	7.76e+01
36	88	3.2	513	9	R47264	Pre-pro BMP6.	6.65e+01
37	88	3.2	513	2	R07311	Human Bone Morphogene	6.65e+01
38	88	3.2	513	6	R32904	Human BMP-6 peptide s	6.65e+01
39	88	3.2	513	7	R36735	Human BMP-6.	6.65e+01
40	87	3.2	542	16	R90772	Bacillus stearothermo	7.76e+01
41	88	3.2	896	17	R92750	Human EGF receptor su	6.65e+01
42	86	3.2	2332	8	R43257	Human Factor VIII.	9.04e+01
43	87	3.2	2351	3	P50319	Human antihaemophilic	7.76e+01
44	86	3.2	2351	3	P50059	Human factor VIII.	9.04e+01
45	86	3.2	2351	13	R78223	Human Factor-VIII:c.	9.04e+01

ALIGNMENTS

RESULT 1
ID R73991 standard; Protein; 537 AA.
AC R73991;
DT 10-JAN-1996 (first entry)
DE Bovine oviduct specific glycoprotein.
KW Bovine oviduct specific glycoprotein; recombinant production;
KW BOGP.
OS Bos taurus.
FH Key Location/Qualifiers
FT Peptide 1..18
FT /label= sig_peptide
FT Peptide 19..537
FT /label= mat_peptide
PN J07107979-A.
PD 25-APR-1995.
PF 15-AUG-1994; 214227.
PR 19-AUG-1993; JP-227881.
PA (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR WPI; 95-190179/25.
DR N-PSDB; Q90442.
PT New DNA encoding an oviduct-specific glycoprotein - useful for
PT recombinant protein production in high quantities.
PS Claim 4; Pages 7-9; 22pp; Japanese.
CC Q90442 encodes R73991 bovine oviduct specific glycoprotein (BOGP).
CC The cDNA can be used for the commercial recombinant prodn. of
CC BOGP in high quantities.
SQ Sequence 537 AA;

Query Match 48.8%; Score 1328; DB 14; Length 537;
Best Local Similarity 51.2%; Pred. No. 7.85e-114;
Matches 187; Conservative 74; Mismatches 97; Indels 7; Gaps 6;

```
Db     20 klvcyftnwafsrpgpasilprdldpflcthlvfafasmsnnqivpkdpqdekilypefn 79
          ||||||||||  | | | :||:|||| |||||::|||:|:|:|:   :  ||  || |||
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDET-LYQEFN 60

Db     80 klkernrglktllsiggwnfgtvrfttmlstfsnrerfvssviallrthgfdgldlffly 139
             ||  |  |||||:|||||||| :|| |::| :||: ||:| | :||  :||||||   |
Qy     61 GLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEY 120

Db    140 pglrgspardrwtfvflleellqafkneaqltmrprlllsaavsgdphvvqkayearllg 199
             || :|||| |:  |  |:::|  ||  ||| : : |||||||||::    |: :||   ::
Qy    121 PGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIA 180
```

```
Db    200 rlldfisvlsydlhgswekvtghnsplfs-l--pgdpks-sa-yamnywrqlgvppekll 254
          : |||:::::||:||||||||||||||||:     :| : | :   |:  | | | |: ||:
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    255 mglptygrtfhllkasqnelraqavgpaspgkytkqagflayyeiccfvrrakkrwindq 314
          :|:|||||:| |  :|:  : | | |:::|| :||::| |||||:|   : | |::| ||
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQ 299

Db    315 yvpyafkgkewvgyddaisfgykaffikrehfggamvwtldlddfrgyfcgtgpfplvht 374
            ||| |: ::|||:||  || |  ::|:  :||||||:|||||| |: |  | :||::|
Qy    300 KVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYPLIQT 359

Db    375 lnnll 379
          |   |
Qy    360 LRQEL 364


RESULT    2
ID   R73993 standard; Protein; 668 AA.
AC   R73993;
DT   10-JAN-1996  (first entry)
DE   Hamster oviduct specific glycoprotein.
KW   Hamster oviduct specific glycoprotein; recombinant production;
KW   HOGP.
OS   Cricetulus griseus.
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT   /label= sig_peptide
FT   Peptide         19..668
FT   /label= mat_peptide
PN   J07107979-A.
PD   25-APR-1995.
PF   15-AUG-1994; 214227.
PR   19-AUG-1993; JP-227881.
PA   (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR   WPI; 95-190179/25.
DR   N-PSDB; Q90444.
PT   New DNA encoding an oviduct-specific glycoprotein - useful for
PT   recombinant protein production in high quantities.
PS   Claim 4; Pages 16-18; 22pp; Japanese.
CC   Q90444 encodes R73993 hamster oviduct specific glycoprotein (HOGP).
CC   The cDNA can be used for the commercial recombinant prodn. of
CC   HOGP in high quantities.
SQ   Sequence   668 AA;

  Query Match           47.9%;  Score 1303;  DB 14;  Length 668;
  Best Local Similarity  48.4%;  Pred. No. 2.08e-111;
  Matches   177;  Conservative   81;  Mismatches  99;  Indels   9;  Gaps   7;

Db     20 klvcyftnwahsrpvpasilprdldpflcthlifafasmsnnqivannlqdekilypefn 79
          ||||||||||: |   | :||:|||| ||||||:|||:|:|:|: : :  ||  || |||
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDET-LYQEFN 60

Db     80 klkernralktllsvggwnfgtsrfttmlstlasrekfigsvvsflrthgfdgldlffly 139
           ||  |  |||||::|||||| :|| |::|  :|: |: | : |||  :||||||   |
Qy     61 GLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEY 120

Db    140 pglrgspindrwnflflieelqfafekealltqrprlllsaav-sgipyiiqtsydvhll 198
          || :|||  |:  |  |:::|  ||: ||  : : |||||||| :|  |: :::|:|  :
Qy    121 PGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYV-DAGYEVDKI 179

Db    199 grrldfinvlsydlhgswekstghnsplfslpedpkssa-famn--y--wrnlgapadkl 253
          :: |||:|:::||:|||||| |||||||:   |:: ::| : ::     |   |:|| ||
Qy    180 AQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKL 239

Db    254 lmgfpaygrtfhllresknglqaasmgpaspgkytkqagflayyevcsfiqraekhwidh 313
          ::| |:|||:| |   |   : |:: |:::|| :||::| ||||||||    | |::|:
Qy    240 ILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQD 298

Db    314 qyvpyaykgkewvgyddavsfsykamfvkkehfggamvwtldmddvrgtfcgngpfplvh 373
          | ||| :: ::|||:||  || |  ::|   :||||||:||:||  |  |  | :||::
Qy    299 QKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYPLIQ 358

Db    374 ilnell 379
            | : |
Qy    359 TLRQEL 364


RESULT    3
ID   R73992 standard; Protein; 718 AA.
AC   R73992;
DT   10-JAN-1996  (first entry)
DE   Murine oviduct specific glycoprotein.
KW   Murine oviduct specific glycoprotein; recombinant production;
KW   MOGP.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT   /label= sig_peptide
FT   Peptide         19..718
FT   /label= mat_peptide
PN   J07107979-A.
PD   25-APR-1995.
PF   15-AUG-1994; 214227.
PR   19-AUG-1993; JP-227881.
PA   (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR   WPI; 95-190179/25.
DR   N-PSDB; Q90443.
PT   New DNA encoding an oviduct-specific glycoprotein - useful for
PT   recombinant protein production in high quantities.
PS   Claim 4; Pages 11-14; 22pp; Japanese.
CC   Q90443 encodes R73992 murine oviduct specific glycoprotein (MOGP).
CC   The cDNA can be used for the commercial recombinant prodn. of
CC   MOGP in high quantities.
SQ   Sequence   718 AA;

  Query Match           47.9%;  Score 1303;  DB 14;  Length 718;
  Best Local Similarity  49.9%;  Pred. No. 2.08e-111;
  Matches   182;  Conservative   76;  Mismatches 100;  Indels   7;  Gaps   4;

Db     20 klvcyftnwahsrpgpasimphdldpflcthlifafasmsnnqivaknlqdenvlypefn 79
          ||||||||||: | | | ::| |||| ||||||:|||:|:|:|: : :  ||  || |||
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDET-LYQEFN 60

Db     80 klkernrelktllsiggwnfgtsrftamlstlanrekfidsvisflrihgfdgldlffly 139
            ||  |  |||||:|||||||| :|| |::|  ||: |::| | |||  :||||||   |
Qy     61 GLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEY 120

Db    140 pglrgspphdrwnflflieelqfaferealltqhprlllsaavsgipsiihtsydalllg 199
          || :|||: |:  |  |:::|  ||::||  :   ||||||||::  : : ::|:   ::
Qy    121 PGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIA 180

Db    200 rrldfinvlsydlhgswekftghnsplfslpedskssa--y---amnywrklgtpadkli 254
          : |||:|:::||:|||||| |||||||:   |:| ::|      |:  |   |||| |||
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    255 mgfptygrnfyllkesknglqtasmgpaspgkytkqagflayyevcsfvqrakkhwidyq 314
          :| |||||:| |    |   : ::: |:::|| :||::| ||||||||    | |::|: |
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQ 299

Db    315 yvpyafkgkewlgyddtisfsykamyvkrehfggamvwtldmddvrgtfcgngpfplvhi 374
            ||| |: ::|:|:||  ||  |  |:|:  :||||||:||:||  |  |  | :||::
Qy    300 KVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYPLIQT 359

Db    375 lnell 379
          | : |
Qy    360 LRQEL 364


RESULT    4
ID   P81342 standard; protein; 321 AA.
AC   P81342;
DT   19-OCT-1990  (first entry)
DE   Polypeptide involved in protective mechanisms
KW   Immune response; cell growth.
PN   J63032898-A.
```

```
PD   30-JAN-1988.
PF   16-JUL-1986; 167518.
PR   16-JUL-1986; JP-167518.
PA   (DAIN) Dainippon Pharm KK.
DR   WPI; 88-068419/10.
DR   P-PSDB; P81342.
PT   New polypeptide and DNA encoding it - related to protective
PT   mechanisms such as immune response etc.
PS   Disclosure; p; Japanese.
CC   This polypeptide is involved in protective mechanisms such as immune
CC   response, cell growth and activation of protective functions.
SQ   Sequence   321 AA;

  Query Match           32.4%;  Score 881;  DB 1;  Length 321;
  Best Local Similarity  50.2%;  Pred. No. 9.90e-71;
  Matches   108;  Conservative   60;  Mismatches  41;  Indels   6;  Gaps   3;

Db     22 klvcyytswsqyregdgscfpdaldrflcthiiysfanisndhidtwewndvtlygmlnt 81
          |||||:|:|:|||:|::  :|  ||  ||||:||:|| ::| :: | |||| |||  :|
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDETLYQEFNG 61

Db     82 lknrnpnlktllsvggwnfgsqrfskiasntqsrrtfiksvppflrthgfdgvdlawlyp 141
          ||: ||:|||||::||||||:|:|: : : : :|:||::|   |||  :|||:|| | ||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    142 g-rg----dkqhfttlikemkaefikeaq-pgkkqlllsaalsagkvtidssydiakisq 195
          | :|    ||::|||| ::   |  ||| :|| :|||||::||   :|::|:: ||:|
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQ 181

Db    196 hldlisimtydfhgawrgttghhsplfrgqedasp 230
          :||::::|:|||||:|   |||:|||:: ||::::
Qy    182 NLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGA 216


RESULT    5
ID   R70025 standard; Protein; 371 AA.
AC   R70025;
DT   29-SEP-1995  (first entry)
DE   Tobacco chitinase encoded by tobacco Cluster-A cDNA clone cA-3.
KW   Tobacco; chitinase; antifungal; fungicide; Cluster-A.
OS   Nicotiana tabacum.
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT   /label= signal peptide
FT   Misc_difference 22
FT   /note= "not present in genomic clone"
FT   Misc_difference 182..195
FT   /note= "Determined by protein sequencing"
FT   Misc_difference 216..241
FT   /note= "Determined by protein sequencing"
PN   EP-639642-A.
PD   22-FEB-1995.
PF   17-AUG-1993; 202425.
PR   17-AUG-1993; EP-202425.
PA   (MOGE-) MOGEN INT NV.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PI   Apotheker-de Groot M,  Bol JF,  Cornelissen BJC,  Linthorst HJM;
PI   Melchers LS,  Ponstein AS,  Sela-buurlage MB;
DR   WPI; 95-083454/12.
DR   N-PSDB; Q82973.
PT   New plant protein having endo-chitinase activity - used in
PT   antifungal compsns. and to develop transformed plants which are
PT   less susceptible to fungal infection.
PS   Claim 3; Page 19-20; 43pp; English.
CC   Screening of a lambda ZAP cDNA library of TMV-infected Samsunn NN
CC   tobacco plants with a probe derived from PROB40 (a partial Cluster-A
CC   cDNA clone) resulted in the isolated of 11 positively hybridising
CC   clones. Analysis revealed that all were identical and corresp. to
CC   Cluster-A cDNA. The nt sequence of clone cA-3 is given in Q82973/
CC   R70025. This clone lacks 7 codons of the 5' part of the ORF. The
CC   cDNA clone was completed by performing a PCR reaction on clone cA-3
CC   with primer T7 (Q82794) and primer P1 (Q82975). A partial AA
CC   sequence of the protein isolated from TMV-infected tobacco leaves
CC   was obtd. and found to be almost identical to the AA sequence
CC   deduced from the cDNA.
SQ   Sequence   371 AA;

  Query Match           21.9%;  Score 597;  DB 13;  Length 371;
  Best Local Similarity  33.0%;  Pred. No. 9.86e-44;
  Matches   107;  Conservative   86;  Mismatches 108;  Indels  23;  Gaps  18;

Db     36 nnidstlfthlfcafadlnpqsnqliispenqds-fsqftstvqrknpsvktflsiaggr 94
          :::|::| |||: ||| :   ::||  :  |::: : :| : : ::|| :||:|:| ||:
Qy     23 KDLDPSLCTHLIYAFAGMT--NHQLSTTEWNDETLYQEFNG-LKKMNPKLKTLLAI-GGW 78

Db     95 a-dttaygimarqpnsrksfidssirlarqfgfhgldldweyp--l-s-atdmtnlgill 149
            |  :  |   :|:| :|::|:||: | ::| |||||||||    | | |:  :  |:
Qy     79 NFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLV 138

Db    150 newrtainmearnsgraallltaavsysprvnglnypvesvarnlnwinlmaydfygpnw 209
           :   |:  ||: ||:  |||:|||: :       | |: :|:||: :|||||| |: |
Qy    139 QDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGS-W 197

Db    210 spsqtnshaqlfdpvnh--isgs---dg-inawiqagvptkklvlgipfygyawrlvnpn 263
             | :: |:   :    ::|   |: :  |:| | |: ||:|||:| || :  | :::
Qy    198 EKV-TGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSS 256

Db    264 ihdlrapaagksnvga-vddgsm-tynrirdyivqtrattvynativgdycysgsnwisy 321
           : |||:| :  |:   :|:| :|  :      |:   : :   |  | : : |:::
Qy    257 DTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQR-IQDQK-VP-YIFRDNQWVGF 313

Db    322 ddtqsvrnkvnyvkgrgllgyfaw 345
           || :| : ||:|:| :|| |   |
Qy    314 DDVESFKTKVSYLKQKGLGGAMVW 337


RESULT    6
ID   R70029 standard; Protein; 377 AA.
AC   R70029;
DT   29-SEP-1995  (first entry)
DE   Tobacco cluster-A protein encoded by genomic clone.
KW   Tobacco; chitinase; antifungal; fungicide; Cluster-A.
OS   Nicotiana tabacum.
FH   Key             Location/Qualifiers
FT   Peptide         1..25
FT   /label= signal peptide
FT   Misc_difference 28..29
FT   /note= "cDNA encodes additional Ile here"
FT   Misc_difference 188..201
FT   /label= Determined by sequencing the protein
FT   Misc_difference 222..247
FT   /note= "Determined by sequencing the protein"
PN   EP-639642-A.
PD   22-FEB-1995.
PF   17-AUG-1993; 202425.
PR   17-AUG-1993; EP-202425.
PA   (MOGE-) MOGEN INT NV.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PI   Apotheker-de Groot M,  Bol JF,  Cornelissen BJC,  Linthorst HJM;
PI   Melchers LS,  Ponstein AS,  Sela-buurlage MB;
DR   WPI; 95-083454/12.
DR   N-PSDB; Q82976.
PT   New plant protein having endo-chitinase activity - used in
PT   antifungal compsns. and to develop transformed plants which are
PT   less susceptible to fungal infection.
PS   Claim 3; Page 25-28; 43pp; English.
CC   Screening of a lambda ZAP cDNA library of TMV-infected Samsunn NN
CC   tobacco plants with a probe derived from PROB40 (a partial Cluster-A
CC   cDNA clone) resulted in the isolated of 11 positively hybridising
CC   clones. Analysis revealed that all were identical and corresp. to
CC   Cluster-A cDNA.The sequence of cDNA clone cA-3 is given in Q82973/
CC   R70025. A genomic library of N. tabacum was screened using the
CC   Cluster-A cDNA insert of clone cA-3 as a probe (see Q82977, Q82978).
CC   The complete nt. sequence of Cluster-A cDNA including the deduced
CC   primary structure of the Cluster-A protein the 5' and 3' UTR regions
CC   of the gene are shown in Q82976/R70029. Comaprison of the cDNA
```

```
CC   clone with the Cluster-A gene revealed that these sequences share
CC   a high degree of identity (94%). The Cluster-A precursor protien
CC   contains a putative signal peptide as well as 4 potential N-linked
CC   glycosylation sites (N-X-S/T). The predicted mature protein has
CC   a calculated mol. wt. of 39,033 Da.
SQ   Sequence   377 AA;

  Query Match           21.9%;  Score 597;  DB 13;  Length 377;
  Best Local Similarity  33.0%;  Pred. No. 9.86e-44;
  Matches   107;  Conservative   88;  Mismatches 106;  Indels  23;  Gaps  18;

Db     42 nnidstlfthlfcafadlnpqlnqliispenqds-frqftstvqrknpsvktflsiaggr 100
          :::|:::| |||: |:| :  :  :|  :  |::: :::| : : ::|| :|:|:|:| ||:
Qy     23 KDLDPSLCTHLIYAFAGMTNH--QLSTTEWNDETLYQEFNG-LKKMNPKLKTLLAI-GGW 78

Db    101 ansta-ygimarqpnsrksfidssirlarqlgfhgldldweyp--l-saa-dmtnlgtll 155
            :|  :  |   :|:| :|:::|:||: |  :| |||||||||    |:| |:  : ||:
Qy     79 NFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLV 138

Db    156 newrtaintearnsgraallltaavsnsprvnglnypveslarnldwinlmaydfygpnw 215
             :   |:  ||: ||:  |||:|||: :       | |: :|:||| :|||||| |: |
Qy    139 QDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGS-W 197

Db    216 spsqtnshaqlfdpvnh--vsgs---dg-inawiqagvptkklvlgipfygyawrlvnan 269
               |  :: |:   :    ::|   |: :  |:| | |: ||:||:| || :  | :::
Qy    198 EKV-TGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSS 256

Db    270 ihglrapaagksnvga-vddgsm-tynrirdyivesrattvynativgdycysgsnwisy 327
             : |||:| :  |:   :|:| :|  :      ::   : :   |  | :  : |:::
Qy    257 DTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQR-IQDQK-VP-YIFRDNQWVGF 313

Db    328 ddtqtvrnkvnyvkgrgllgyfaw 351
             || :: : ||:|:| :|| |   |
Qy    314 DDVESFKTKVSYLKQKGLGGAMVW 337


RESULT    7
ID   R63441 standard; Protein; 424 AA.
AC   R63441;
DT   06-JUL-1995  (first entry)
DE   Trichoderma harzianum P1 endochitinase.
KW   Trichoderma harzianum P1; endochitinase; inhibition;
KW   chitin containing fungi; herbivorous insects.
OS   Trichoderma harzianum.
PN   WO9424288-A.
PD   27-OCT-1994.
PF   10-FEB-1994; U01198.
PR   14-APR-1993; US-045269.
PA   (CORR ) CORNELL RES FOUND INC.
PI   Harman GE,  Hayes CK,  Lorito M,  Tronsmo A;
DR   WPI; 94-341867/42.
DR   N-PSDB; Q78144.
PT   Trichoderma harzianum P1 endo:chitinase and chitobiase genes -
PT   useful in biological control of chitin-contg. fungi and insects
PT   (Eng).
PS   Example 8; Page 45; 62pp; English.
CC   Q78144 encodes R63441 the Trichoderma harzianum P1 endochitinase,
CC   which can be used to inhibit chitin containing fungi and chitin
CC   containing herbivorous insects, especially from the genera below
CC   Fusarium, Gliocladium, Rhizoctonia, Trichoderma, Uncinula, Ustilage,
CC   Erysiphe, Botrytis, Saccharomyces, Sclerotium and Alternaria.
SQ   Sequence   424 AA;

  Query Match           21.4%;  Score 581;  DB 12;  Length 424;
  Best Local Similarity  36.0%;  Pred. No. 3.16e-42;
  Matches   102;  Conservative   69;  Mismatches  95;  Indels  17;  Gaps  13;

Db    112 qlfk-lkkangnlkvmlsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidv 169
          | |: ||| | :|| :|:|||| :| :|    :|  ||  | ::|| |::  :|||:|:
Qy     57 QEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 116

Db    170 dweyp-addtqatnmvlllkeirsqldayaaqy-apg-yhfllsiaapagpehy-sflhm 225
          ||||| :: : | ::  :   ::   :|:  :  ::|  ::||| | |||     :   :

Qy    117 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 176

Db    226 sdlgqvldyvnlmaydyagswssysghdanlfanpsnpnssp-yntdqaikdyikggvpa 284
             ::| ||:|||||||: |||   :||:: |:    :: :::  | |~|: : :  | ||
Qy    177 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 236

Db    285 skivlgmpiygraf--estgg--igqtysgigs-gsw--engiwdy-kvlpkagatvqyd 336
          ||::|||| |||:|   |::   :|   :| |: |:   | |:  |  | :  ||| |
Qy    237 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQ-- 294

Db    337 svaqayysydpsskelisfdtpdmintkvsylknlglggsmfw 379
             : :   :|    :: ::||  : ::|||||||  ||||:| |
Qy    295 RIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT    8
ID   R56861 standard; Protein; 389 AA.
AC   R56861;
DT   17-FEB-1995  (first entry)
DE   Aphanocladium album mature endochitinase.
KW   oxalic acid degradation; oxalate oxidase; selection marker;
KW   filamentous fungus endochitinase; precursor;
KW   pathogen resistance; plant cell selection.
OS   Aphanocladium album.
FH   Key             Location/Qualifiers
FT   mat_peptide     1..1167
FT   /*tag= a
FT   /product= endochitinase
PN   WO9413790-A.
PD   23-JUN-1994.
PF   07-DEC-1993; F01203.
PR   07-DEC-1992; FR-014721.
PA   (SNFI ) ELF SANOFI.
PA   (ERAP ) SOC NAT ELF AQUITAINE.
PI   Grezes-Besset B,  Grison R,  Pignard A,  Schneider M;
DR   WPI; 94-217870/26.
DR   N-PSDB; Q68353.
PT   Use of gene encoding oxalic acid degrading enzyme - for plant
PT   cell selection, esp. gene coupled to gene-encoding protein
PT   conferring pathogen resistance
PS   Disclosure and Claim 8; Page 39-40; 64pp; French.
CC   The sequence coding for wheat germin, a protein induced during
CC   germination of wheat, is a preferred example of DNA coding for a
CC   protein able to degrade oxalic acid. Such DNA can be coupled to a
CC   sequence of interest, e.g. one coding for enzymes which confer
CC   resistance to pathogens (esp. endochitinases). Plant calli which
CC   have been successfully transformed by the sequences are selected
CC   by their ability to grow on oxalic acid-contg. medium. R56861 is
CC   a known sequence of a mature endochitinase; the sequence was
CC   previously disclosed in EP-531218.
SQ   Sequence   389 AA;

  Query Match           21.2%;  Score 576;  DB 10;  Length 389;
  Best Local Similarity  36.2%;  Pred. No. 9.31e-42;
  Matches   101;  Conservative   65;  Mismatches  95;  Indels  18;  Gaps  16;

Db     81 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 139
          ||| | ::| :|:|||| :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    140 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 194
           :::: | ::  :   |:  :::  :  : |  ::||| | ||| : |    :  : ::
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 180

Db    195 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 253
             ||::||||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    254 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 305
          |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 300

Db    306 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 344
```

```
          | :    :: ::||    : |||| || |||||:| |
Qy    301 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT    9
ID   R33072 standard; Protein; 423 AA.
AC   R33072;
DT   30-JUN-1993  (first entry)
DE   Aphanocladium album pre-pro-endochitinase.
KW   filamentous fungus; pathogen resistance.
OS   Aphanocladium album.
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT   /label= signal
FT   /note= "pre-sequence"
FT   Peptide         23..34
FT   /note= "pro-sequence"
FT   Protein         35..423
FT   /label= endochitinase
PN   EP-531218-A.
PD   10-MAR-1993.
PF   04-SEP-1992; 402414.
PR   06-SEP-1991; FR-011072.
PA   (SNFI ) ELF SANOFI.
PA   (ERAP ) SOC NAT ELF AQUITAINE.
PI   Blaiseau P,  Legoux R,  Leguay J,  Schneider M;
DR   WPI; 93-078700/10.
DR   N-PSDB; Q37711, Q37712.
PT   Recombinant DNA encoding fungal endo:chitinase - useful for
PT   protecting plants against fungi, insects, etc.
PS   Claim 1; Fig 1; 75pp; French.
CC   Clone CH3C was isolated from a cDNA bank prepared from A.album
CC   mRNA by screening with anti-chitinase antiserum. Three probes were
CC   designed based on the CH3C sequence, corresponding to the start,
CC   middle and downstream regions of the coding sequence. The probes
CC   were used to isolate the full-length coding sequence of A.album
CC   endochitinase from an A.album cDNA library in pTZ19R. A 1.6kb
CC   fragment was identified in several positive clones. The sequence
CC   of the fragment contained an ORF encoding a 423 amino acid protein.
CC   The mature endochitinase can be used to control pathogens
CC   such as fungi, bacteria, arthropods and nematodes on plants.
SQ   Sequence   423 AA;

  Query Match           21.2%;  Score 576;  DB 6;  Length 423;
  Best Local Similarity  36.2%;  Pred. No. 9.31e-42;
  Matches   101;  Conservative   65;  Mismatches  95;  Indels  18;  Gaps  16;

Db    115 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 173
          ||| | ::| :|:|||| :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    174 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 228
           :::: | ::  :   |:   ::: :  : |  ::||| | ||| : |    :  : ::
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 180

Db    229 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 287
          ||::|||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    288 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 339
          |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 300

Db    340 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 378
          | :    :: ::||    : |||| || |||||:| |
Qy    301 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT   10
ID   R33068 standard; Protein; 389 AA.
AC   R33068;
DT   30-JUN-1993  (first entry)
DE   Aphanocladium album mature endochitinase.
KW   filamentous fungus; pathogen resistance.
OS   Aphanocladium album.
PN   EP-531218-A.
PD   10-MAR-1993.
PF   04-SEP-1992; 402414.
PR   06-SEP-1991; FR-011072.
PA   (SNFI ) ELF SANOFI.
PA   (ERAP ) SOC NAT ELF AQUITAINE.
PI   Blaiseau P,  Legoux R,  Leguay J,  Schneider M;
DR   WPI; 93-078700/10.
DR   N-PSDB; Q37704.
PT   Recombinant DNA encoding fungal endo:chitinase - useful for
PT   protecting plants against fungi, insects, etc.
PS   Claim 22; Page 60-61; 75pp; French.
CC   This mature protein from the filamentous fungus Aphanocladium
CC   album has endochitinase activity and is used to control pathogens
CC   such as fungi, bacteria, arthropods and nematodes on plants.
CC   Recombinant DNA comprising the endochitinase coding sequence is
CC   specifically intended for transforming Nicotiana tabacum, Helianthus
CC   annuus and Brassica napus to pathogen resistance.
SQ   Sequence   389 AA;

  Query Match           20.9%;  Score 569;  DB 6;  Length 389;
  Best Local Similarity  35.8%;  Pred. No. 4.23e-41;
  Matches   100;  Conservative   66;  Mismatches  95;  Indels  18;  Gaps  16;

Db     81 lkkqnrnmkvmvsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 139
          ||| | ::| :::|||| :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    140 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 194
           :::: | ::  :   |:   ::: :  : |  ::||| | ||| : |    :  : ::
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 180

Db    195 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 253
          ||::|||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    254 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 305
          |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 300

Db    306 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 344
          | :    :: ::||    : |||| || |||||:| |
Qy    301 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT   11
ID   W02159 standard; Protein; 866 AA.
AC   W02159;
DT   14-JAN-1997  (first entry)
DE   Soluble chitinase.
KW   Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;
KW   Beta-N-acetylglucosamidase; chitin; oligosaccharide; catabolic;
KW   catabolism.
OS   Vibrio furnissii.
PN   WO9625424-A1.
PD   22-AUG-1996.
PF   13-FEB-1996; U02332.
PR   13-FEB-1995; US-386727.
PA   (UYJO ) UNIV JOHNS HOPKINS.
PI   Bassler B,  Chitlaru E,  Keyhani N,  Roseman S,  Rowe C;
PI   Yu C;
DR   WPI; 96-393335/39.
DR   N-PSDB; T36390.
PT   Chitin biosynthetic enzymes end I, exo I and exo II  - are
PT   periplasmic chito:dextrinase(s), periplasmic beta-GlcNAcidase(s) and
PT   aryl beta-N-acetyl:gluco:amidase(s), respectively
PS   Example 4; Page 79-82; 101pp; English.
CC   Periplasmic chitodextrinase (W02156), periplasmic
CC   Beta-N-acetylglucosaminidase (W02157) and aryl
CC   Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin
CC   oligosaccharides with the structure (GlcNAc)n where n is 2 or
```

```
CC   higher, by contacting them with soluble chitin.  The enzymes are
CC   encoded by the genes endI, exoI and exoII respectively.  They are
CC   all genes involved in the catabolic pathway of chitin.
SQ   Sequence   866 AA;

  Query Match           17.0%;  Score 463;  DB 19;  Length 866;
  Best Local Similarity  30.7%;  Pred. No. 3.32e-31;
  Matches    93;  Conservative   81;  Mismatches 104;  Indels  25;  Gaps  20;

Db    253 fgnlmalkkanpnlkilpsvggwtlsdp-fy-ffsdktkrdtfvasmkeylqtwkffdgv 310
          : :: :||| ||:|| | ::||| ::   |    :   :|:||| |   :|:   | ||:
Qy     56 YQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSF-DGL 114

Db    311 didwefpggqqganpnlggpndgatyvalmkelramldeleaet-grqyeltsaisaggdk 369
          |:|||:||:||: | :   ::   : :|: :|   ::: ||:| |::  | ||  ::|:
Qy    115 DLDWEYPGSQGS-PAVD--KE--RFTTLVQDLANAFQQ-EAQTSGKERLLLSAAVPAGQT 168

Db    370 iakvdyqaaq--qymdyiflmsydfsgafdlknlahqtnly--asswdpatkyttdkgvk 425
              |:     | :|:: ||:||| |: : |  :| : ||       :|:    | :|
Qy    169 YVDAGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNSPLYKRQEESGAAASLNVDAAVQ 227

Db    426 allgqgvtpgkvvvgaamygrgwt-gvng-yqagnpftgsat-gpikgtwengvvdyrdi 482
            |  |  ::|:::| : |||: | : ::  : | | |||:| ||:    | |:: | ::
Qy    228 QWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFT-K-EGGMLAYYEV 285

Db    483 vnnrmgagweqggydetaeapyvfkastgdlisfdndrsvkakgqyvlanqlgglfaweid 542
           : ::||   |  ::    ||:|: :  : ::||:  | |:|  |:  : |||   | :|
Qy    286 CS-WKGAT-KQRIQD-QKVPYIFRDN--QWVGFDDVESFKTKVSYLKQKGLGGAMVWALD 340

Db    543 adn 545
            |:
Qy    341 LDD 343


RESULT   12
ID   R64823 standard; Protein; 561 AA.
AC   R64823;
DT   23-JUL-1995  (first entry)
DE   Chitinase A.
KW   Chitinase A; protein secretion; biological control agent;
KW   transgenic plant; chitin; plant pathogen; nematode; fungus;
KW   crop protection; pCHIT1251.
OS   Serratia marcescens QMB1466.
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT   /label= Sig_peptide
PN   US5374540-A.
PD   20-DEC-1994.
PF   26-MAR-1984; 593691.
PR   26-MAR-1984; US-593691.
PR   18-JUL-1986; US-888033.
PR   09-JUL-1990; US-550253.
PA   (DNAP ) DNA PLANT TECHNOLOGY CORP.
PI   Jones JDG,  Suslow T;
DR   WPI; 95-035648/05.
DR   N-PSDB; Q76290.
PT   Plant cell contg. signal sequence of chitinase A gene - for
PT   secretion of foreign polypeptide, esp. chitinase A for protection
PT   against pathogenic fungi and nematodes.
PS   Disclosure; Fig. 1A-1C; 20pp; English.
CC   Serratia marcescens QMB1466 genomic DNA was partially digested,
CC   ligated into vector pLAFR1, packaged into lambda phage particles,
CC   and used to transfect Escherichia coli DH1 (ATCC 33849) cells.
CC   Plating on chitin medium was used to screen for chitinase activity.
CC   Clone C12 (ATCC 67152) was isolated.  Its sequence is given in
CC   Q76290, and the deduced protein sequence in R64823.
SQ   Sequence   561 AA;

  Query Match           14.3%;  Score 389;  DB 12;  Length 561;
  Best Local Similarity  28.2%;  Pred. No. 2.18e-24;
  Matches    91;  Conservative   81;  Mismatches 126;  Indels  25;  Gaps  18;

Db    231 pfaalqkaqkgvtawddpykgnfgqlmalkqahpdlkilpsiggwtlsdpfff-mgdkvk 289
          :||:: : | : | |:|     : :: :||  :| || | :|||| ::   |  |    :
Qy     36 AFAGMTNHQLSTTEWNDE-T-LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATAN 93

Db    290 -rdrfvgsvkeflqtwkffdgvdidwefpggkganpnlgspqdgetyvllmkelramldq 348
           |: || |   ||:   | ||:|:|||:||: |: | :    | | :  |: :|   ::|
Qy     94 NRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV----DKERFTTLVQDLANAFQQ 147

Db    349 lsaetgr-kyeltsaisagkdkidkvaynvaq--nsmdhiflmsydfygafdlknlghqt 405
           :  :|: :  |::|::||   :|  :|:|     ::| : ||:||| |: : |  || :
Qy    148 EAQTSGKERLLLSAAVPAGQTYVD-AGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNS 205

Db    406 alnar---pgsrhrlhhgerrecaagqgvkpgkivvgtamygrgwtgvngyqnnipftgt 462
          :|  |   :|:   |:      :     |  ::|:::| : |||: |  :: :  :   :|
Qy    206 PLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPAT 265

Db    463 hravkgtw--engivdyrqiasqfmsgewqytydataeapyvfkpstgdlitsddarsvq 520
            :  |    | |:: | :: |  :::  |   :     ||:|:    : :  ||  |
Qy    266 GSGTPGPFTKEGGMLAYYEVCS-WKGATKQ-RIQD-QKVPYIFR-DN-QWVGFDDVESFK 320

Db    521 akgkyvldkqlgglfsweidadn 543
          :|  |: :| |||    | :| |:
Qy    321 TKVSYLKQKGLGGAMVWALDLDD 343


RESULT   13
ID   R48981 standard; Protein; 561 AA.
AC   R48981;
DT   04-SEP-1994  (first entry)
DE   Bacterial chitinase-A fragment from pCHIT1251.
KW   Chitinase-A; enzyme; transgenic plant; biological control;
KW   disease-resistance; crop improvement; Agrobacterium; fungus;
KW   nematode; insect.
PN   US5290687-A.
PD   01-MAR-1994.
PF   26-MAR-1984; 593691.
PR   26-MAR-1984; US-593691.
PR   18-JUL-1986; US-888033.
PR   09-JUL-1990; US-550253.
PR   14-AUG-1992; US-930970.
PA   (DNAP ) DNA PLANT TECHNOLOGY CORP.
PI   Jones JDG,  Suslow T;
DR   WPI; 94-074348/09.
DR   N-PSDB; Q57966.
PT   Inhibition of chitinous fungal plant pathogens - by transforming
PT   plant with DNA encoding chitinase activity
PS   Disclosure; Fig.1; 13pp; English.
CC   Plasmid pCHIT1251 containing the chitinase-A gene (chiA), fused
CC   directly to a nos promoter,  may be used to transform, via
CC   Agrobacterium sp., a crop plant.  The resulting transgenic plants
CC   expressing the chitinase-A protein fragment are protected against
CC   fungi, nematodes, insects and disease agents. The protein is
CC   preferably expressed by the plant only at the site of pathogen
CC   attack, e.g. only in the root cells.  The plants express the
CC   chitinase in a biologically active form.
SQ   Sequence   561 AA;

  Query Match           13.7%;  Score 372;  DB 9;  Length 561;
  Best Local Similarity  27.6%;  Pred. No. 7.79e-23;
  Matches    89;  Conservative   84;  Mismatches 125;  Indels  25;  Gaps  18;

Db    231 pfaalqkaqkgvtawddpykgnfgqlmalkqahpdlkilpsiggwtlsdpfff-mgdkvk 289
          :||:: : | : | |:|     : :: :||  :| || | :|||| ::   |  |    :
Qy     36 AFAGMTNHQLSTTEWNDE-T-LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATAN 93

Db    290 -rdrfvgsvkeflqswkffdgvdidwefpggkganpnlgspqdgetyvlvmkevgamldq 348
           |: || |   ||:   | ||:|:|||:||: |: | :    | | :  :: :::  ::|
Qy     94 NRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV----DKERFTTLVQDLANAFQQ 147

Db    349 lwaetgr-kyeltsaisagkdkidkvaynvaq--nsmdhiflmsydfygafdlknlghqt 405
              :|: :  |::|::||   :|  :|:|     ::| : ||:||| |: : |  || :
Qy    148 EAQTSGKERLLLSAAVPAGQTYVD-AGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNS 205

Db    406 alnar---pgsrhrlhhgegrecaagqgvkpgkivvgtamygrgwtgvngyqnnipftgt 462
```

```
          :|  |   :|:   |:   : :     |  ::|:::| : |||: |  :: :  :   :|
Qy    206 PLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPAT 265

Db    463 hravkgtw--engivdyrqiasqfmsgewqytydgtaeapyvfkpstgdlitsddarsvq 520
          :  |    | |:: | :: |  :::  |   :     ||:|:    : :  ||  |
Qy    266 GSGTPGPFTKEGGMLAYYEVCS-WKGATKQ-RIQD-QKVPYIFR-DN-QWVGFDDVESFK 320

Db    521 akgkyvldkqlgglfsweidadn 543
          :|  |: :| |||    | :| |:
Qy    321 TKVSYLKQKGLGGAMVWALDLDD 343


RESULT   14
ID   R32547 standard; Protein; 238 AA.
AC   R32547;
DT   17-JUN-1993  (first entry)
DE   HaNPV IE-1 gene protein (partial sequence).
KW   Heliothis armigera nuclear polyhedrosis virus; Ha; NPV;
KW   insecticide; immediate early-1.
OS   Heliothis armigera nuclear polyhedrosis virus.
PN   WO9303144-A.
PD   18-FEB-1993.
PF   05-AUG-1992; AU0413.
PR   05-AUG-1991; AU-007576.
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PI   Christian PD, Dall DJ, Gordon KHJ, Hanzlik TN, Sriskantha A.
DR   WPI; 93-076501/09.
DR   N-PSDB; Q35982.
PT   Insect virus with reduced capacity to occlude viral particles - used
PT   for controlling proliferation of insect pests without horizontal
PT   transmission
PS   Disclosure; Fig 2; 51pp; English.
CC   The sequence is that encoded by the Heliothis armigera (Ha)
CC   nuclear polyhedrosis virus (NPV) immediate early-1 (IE-1) gene
CC   (partial sequence).
SQ   Sequence   238 AA;

  Query Match           10.0%;  Score 271;  DB 6;  Length 238;
  Best Local Similarity  34.1%;  Pred. No. 9.19e-14;
  Matches    58;  Conservative   43;  Mismatches  51;  Indels  18;  Gaps  14;

Db     44 erdkflsnpgwrefvltwkf-fdgididwefpggkganpnvgdverdnntyiallgelra 102
          :|: |::::  | | | :|: |||:|:|||:||: |: | | | ||    : :|: :|
Qy     94 NRQTFVNSAI-R-F-L-RKYSFDGLDLDWEYPGSQGS-PAV-DKER----FTTLVQDLAN 143

Db    103 mldq-vqiqtnrtlelttaisagidkiaav-ncdraqqyldkifvmsydfkgawsntdlg 160
           ::|  |   :  | |::|::||   : |  : |:  | || : :|:||| |:| :   |
Qy    144 AFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVT-G 202

Db    161 hqtaly--gsawkpnepytanvavdallaqrvnp-kklvlgvamygrgwt 207
          | ::||       :  :   : ||:  | |:  |  ||:||:: |||: |
Qy    203 HNSPLYKRQEESGAAASLNVDAAVQQWL-QKGTPASKLILGMPTYGRSFT 251


RESULT   15
ID   W02156 standard; Protein; 1046 AA.
AC   W02156;
DT   14-JAN-1997  (first entry)
DE   Periplasmic chitodextrinase.
KW   Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;
KW   Beta-N-acetylglucosamidase; chitin; oligosaccharide; catabolic;
KW   catabolism.
OS   Vibrio furnissii.
PN   WO9625424-A1.
PD   22-AUG-1996.
PF   13-FEB-1996; U02332.
PR   13-FEB-1995; US-386727.
PA   (UYJO ) UNIV JOHNS HOPKINS.
PI   Bassler B,  Chitlaru E,  Keyhani N,  Roseman S,  Rowe C;
PI   Yu C;
DR   WPI; 96-393335/39.
DR   N-PSDB; T36387.
PT   Chitin biosynthetic enzymes end I, exo I and exo II  - are
PT   periplasmic chito:dextrinase(s), periplasmic beta-GlcNAcidase(s) and
PT   aryl beta-N-acetyl:gluco:amidase(s), respectively
PS   Claim 2; Page 68-71; 101pp; English.
CC   Periplasmic chitodextrinase (W02156), periplasmic
CC   Beta-N-acetylglucosaminidase (W02157) and aryl
CC   Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin
CC   oligosaccharides with the structure (GlcNAc)n where n is 2 or
CC   higher, by contacting them with soluble chitin.  The enzymes are
CC   encoded by the genes endI, exoI and exoII respectively.  They are
CC   all genes involved in the catabolic pathway of chitin.
SQ   Sequence   1046 AA;

  Query Match           8.0%;  Score 219;  DB 19;  Length 1046;
  Best Local Similarity  33.3%;  Pred. No. 3.07e-09;
  Matches    48;  Conservative   40;  Mismatches  42;  Indels  14;  Gaps  10;

Db    454 etfadsavemmrkyrfdgldidlrisnidggtgnpddtafsesrraylmnsyhelmrvlr 513
          :|| :||:  :||| ||||||:| :  :  |: |:|      | | : |:  ::|  ::
Qy     96 QTFVNSAIRFLRKYSFDGLDLD-W--EYPGSQGSPA-VD-KE-RFTTLV---QDLANAFQ 146

Db    514 ekldvasaqdgvhymltiaapss-ayllrgmetmavtqyldyvnimsydlhgawndhvgh 572
          :  : :|: :  : :|: | |:: :|:  | |   ::| ||:||:|:||:||:|:   ||
Qy    147 QEAQ-TSGKE--RLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGH 203

Db    573 naalydtgkdselaqw-nvygtaq 595
          |:::|    :|  |   || :: |
Qy    204 NSPLYKRQEESGAAASLNVDAAVQ 227


Search completed: Fri Jun 27 17:04:25 1997
Job time : 74 secs.
```

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 27 17:01:00 1997; MasPar time 17.84 Seconds
596.027 Million cell updates/sec

Tabular output not generated.

Title: >US-08-663-618A-14
Description: (1-373) from US08663618A.pep
Perfect Score: 2721
Sequence: 1 AKLVCYFTNWAQYRQGEARF.....YPLIQTLRQELSLPYLPSGT 373

Scoring table: PAM 150
Gap 11

Searched: 89912 seqs, 28507787 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir50
1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8
13:unann9 14:unann10 15:unenc 16:unrev

Statistics: Mean 47.350; Variance 104.053; scale 0.455

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1488	54.7	383	13	A49562	cartilage glycoprote	8.90e-251
2	1428	52.5	383	16	S51327	heparin-binding glyc	2.76e-239
3	1417	52.1	381	14	S61550	BRP39 protein - mous	3.51e-237
4	1417	52.1	381	14	I48271	BRP39 protein - mous	3.51e-237
5	1345	49.4	399	5	S27879	secretory protein YM	2.04e-223
6	1328	48.8	537	14	S57197	oviduct-specific gly	3.60e-220
7	1318	48.4	539	14	I46470	estrogen dependent o	2.92e-218
8	1179	43.3	483	12	A53918	chitinase (EC 3.2.1.	8.78e-192
9	1086	39.9	504	12	A38221	chitinase (EC 3.2.1.	3.90e-174
10	1015	37.3	554	12	A56596	chitinase (EC 3.2.1.	1.05e-160
11	698	25.7	699	10	A38368	chitinase (EC 3.2.1.	2.18e-101
12	594	21.8	378	5	S51591	chitinase (EC 3.2.1.	2.86e-82
13	582	21.4	424	5	S47133	chitinase (EC 3.2.1.	4.43e-80
14	576	21.2	423	5	JQ1975	chitinase (EC 3.2.1.	5.50e-79
15	561	20.6	423	12	S51369	chitinase - fungus (	2.95e-76
16	546	20.1	427	12	JC4565	chitinase (EC 3.2.1.	1.56e-73
17	446	16.4	563	16	S60651	chitinase precursor	1.45e-55
18	404	14.8	561	9	A25090	chitinase (EC 3.2.1.	3.83e-48
19	345	12.7	799	8	PC4106	chitinase (EC 3.2.1.	6.99e-38
20	344	12.6	499	9	S52422	chitinase (EC 3.2.1.	1.04e-37
21	344	12.6	820	16	A40633	chitinase (EC 3.2.1.	1.04e-37
22	337	12.4	499	9	S04856	chitinase (EC 3.2.1.	1.65e-36
23	314	11.5	511	11	S61166	probable membrane pr	1.37e-32
24	287	10.5	610	10	JH0573	chitinase (EC 3.2.1.	4.76e-28
25	266	9.8	452	12	JC4038	47K glycoprotein - f	1.43e-24
26	263	9.7	1146	11	S07915	RF2 protein - yeast	4.45e-24
27	246	9.0	413	5	JC2135	chitinase (EC 3.2.1.	2.62e-21
28	229	8.4	124	5	S57715	chitinase (EC 3.2.1.	1.40e-18
29	172	6.3	36	14	A27682	39K whey protein - b	7.80e-10
30	172	6.3	83	13	A37954	estradiol-stimulated	7.80e-10
31	136	5.0	29	14	S57204	oviduct-specific gly	8.95e-05
32	136	5.0	29	14	S55693	oviduct-specific sia	8.95e-05
33	121	4.4	597	5	S33848	chitinase (EC 3.2.1.	7.93e-03
34	121	4.4	597	5	S32039	chitinase (EC 3.2.1.	7.93e-03
35	112	4.1	642	16	S37623	homologies with orf	1.02e-01
36	110	4.0	488	10	A41961	chitinase (EC 3.2.1.	1.77e-01
37	104	3.8	305	9	S28440	probable coproporphy	8.88e-01
38	104	3.8	422	4	ZWECC	umuC protein - Esche	8.88e-01
39	100	3.7	364	11	S45925	hypothetical protein	2.52e+00
40	101	3.7	562	3	KRHUEB	keratin, 56K type II	1.94e+00
41	101	3.7	662	11	S61193	probable membrane pr	1.94e+00
42	101	3.7	1045	2	SUSEMM	serine proteinase (E	1.94e+00
43	101	3.7	1070	11	S19686	alpha-glucosidase (E	1.94e+00
44	99	3.6	398	5	A37768	streptococcal pyroge	3.25e+00
45	99	3.6	958	1	JN0102	glucan 1,4-alpha-glu	3.25e+00

ALIGNMENTS

```
RESULT    1
ENTRY           A49562     #type complete
TITLE           cartilage glycoprotein gp39 precursor - human
ALTERNATE_NAMES 39K synovial protein
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change
                  03-May-1996
ACCESSIONS      A49562; S10677; A33162
REFERENCE       A49562
  #authors      Hakala, B.E.; White, C.; Recklies, A.D.
  #journal      J. Biol. Chem. (1993) 268:25803-25810
  #title        Human cartilage gp-39, a major secretory product of articular
                  chondrocytes and synovial cells, is a mammalian member of a
                  chitinase protein family.
  #accession    A49562
    ##status         preliminary
    ##molecule_type mRNA
    ##residues      1-383 ##label HAK
    ##cross-references GB:M80927
REFERENCE       S10677
  #authors      Nyirkos, P.; Golds, E.E.
  #journal      Biochem. J. (1990) 269:265-268
  #title        Human synovial cells secrete a 39 kDa protein similar to a
                  bovine mammary protein expressed during the non-lactating
                  period.
  #cross-references MUID:90328983
  #accession    S10677
    ##molecule_type protein
    ##residues      22-40,'X',42-45 ##label NY2
KEYWORDS        cartilage; extracellular protein; glycoprotein
FEATURE
  1-21               #domain signal sequence #status predicted #label SIG\
  22-383             #product cartilage glycoprotein gp39 #status predicted
                       #label MAT
SUMMARY         #length 383  #molecular-weight 42613  #checksum 2942

  Query Match          54.7%;  Score 1488;  DB 13;  Length 383;
  Best Local Similarity  53.3%;  Pred. No. 8.90e-251;
  Matches  195;  Conservative   83;  Mismatches  79;  Indels   9;  Gaps   5;

Db      23 klvcyytswsqyregdgscfpdaldrflcthiiysfanisndhidtwewndvtlygmlnt 82
           |||||:|:|:|||:|::  :|  ||  ||||:||:|| ::| :: | |||| |||  :|
Qy       2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDETLYQEFNG 61

Db      83 lknrnpnlktllsvggwnfgsqrfskiasntqsrrtfiksvppflrthgfdgldlawlyp 142
```

```
          ||: ||:|||||::||||||:|:|: : : : :|:||::|   |||  :|||||| | ||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    143 grr-----dkqhfttlikemkaefikeaq-pgkkqlllsaalsagkvtidssydiakisq 196
          | :     ||::|||||: ::   |  ||| :|| :|||||||::||   :|::|:: ||:|
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQ 181

Db    197 hldfisimtydfhgawrgttghhsplfrgqedaspdrfsntdyavgymlrlgapasklvm 256
          :|||:::|:|||||:|   |||:|||:: ||:::    | | ||   |: |:|||||::
Qy    182 NLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLIL 241

Db    257 giptfgrsftlass-etgvgapisgpgipgrftkeagtlayyeicdflrgatvhrtlgqq 315
          |:||:|||||||| :| |||| :|:| || ||||:| |||||:|   :||| :|   |
Qy    242 GMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQK 300

Db    316 vpyatkgnqwvgyddqesvkskvqylkdrqlagamvwaldlddfqgsfcgqdlrfpltna 375
          |||  : |||||:|| || |:|| |||:: |:|||||||||||| |  | |  |:||  :
Qy    301 VPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQG-RYPLIQT 359

Db    376 ikdala 381
          ::: |:
Qy    360 LRQELS 365


RESULT    2
ENTRY             S51327     #type complete
TITLE             heparin-binding glycoprotein 38K - pig
ORGANISM          #formal_name Sus scrofa domestica #common_name domestic pig
DATE              07-May-1995 #sequence_revision 07-May-1995 #text_change
                    07-May-1995
ACCESSIONS        S51327
REFERENCE         S51327
   #authors       Shackelton, L.M.; Mann, D.M.; Millis, A.J.T.
   #submission    submitted to the EMBL Data Library, January 1995
   #description   Identification of a 38kDa heparin-binding glycoprotein
                    (gp38k) in differentiating vascular smooth muscle cells as
                    a member of a group of proteins associated with tissue
                    remodeling.
   #accession     S51327
      ##status         preliminary
      ##residues       1-383 ##label SHA
      ##cross-references EMBL:Z47803
SUMMARY           #length 383  #molecular-weight 42443  #checksum 2907

  Query Match           52.5%;  Score 1428;  DB 16;  Length 383;
  Best Local Similarity  52.2%;  Pred. No. 2.76e-239;
  Matches   191;  Conservative   81;  Mismatches  85;  Indels   9;  Gaps   5;

Db     23 klvcyytswsqyregdgscfpdaidpflcthiiysfanisnneidtlewndvtlydtlnt 82
          |||||:|:|:|||:|::  :|  :|| ||||:||:|| ::|::: | |||| |||: :|
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDETLYQEFNG 61

Db     83 lknrnpnlktllsvggwnfgsqrfskiasntqsrrtfiksvppflrthgfdgldlawisp 142
          ||: ||:|||||::||||||:|:|: : : : :|:||::|   |||  :|||||| |  |
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    143 grr-----dkrhlttlvkemkaefvreal-pgterlllsgavsagkvaidrgydiaqisq 196
          | :     || ::|||| ::   | :||  :| ||||||:||:||   :| ||::  |:|
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQ 181

Db    197 hldfislltydfhgawrqttghhsplfrgqgdassdrfsnadyavsyvlrlgapanklvm 256
          :|||::|::|||||:|   |||:|||:: | ::::    | | ||   |: |:||:||::
Qy    182 NLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLIL 241

Db    257 giptfgrsftlassk-tdvgapasgpgipgrftkekgilayyeicdflqgatvrrplgqq 315
          |:||:|||||||||  | |||||:|:| || |||| |:|||||:|    ||| :|   |
Qy    242 GMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQK 300

Db    316 vpyatkgnqwvgyddqesvknkakylksrqlagamvwtldlddfrgnfcgqnlrfpltsa 375
          |||  : |||||:|| || | |  ||| : |:|||||:|||||| |  | |  |:||  :
Qy    301 VPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQG-RYPLIQT 359

Db    376 ikdvla 381
          ::: |:
Qy    360 LRQELS 365


RESULT    3
ENTRY             S61550     #type complete
TITLE             BRP39 protein - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change
                    03-May-1996
ACCESSIONS        S61550
REFERENCE         S61550
   #authors       Morrison, B.W.
   #submission    submitted to the EMBL Data Library, November 1995
   #accession     S61550
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-381 ##label MOR
      ##cross-references EMBL:X93035
SUMMARY           #length 381  #molecular-weight 43001  #checksum 5107

  Query Match           52.1%;  Score 1417;  DB 14;  Length 381;
  Best Local Similarity  50.7%;  Pred. No. 3.51e-237;
  Matches   186;  Conservative   84;  Mismatches  86;  Indels  11;  Gaps   9;

Db     23 klvcyftswsqyregvgsflpdaiqpflcthiiysfanissdnmlstwewndesnydkln 82
          |||||||:|:|||:| : |||  ::| ||||:||:|| :: :: ||| |||||: |: :|
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMT-NHQLSTTEWNDETLYQEFN 60

Db     83 klktrntnlktllsvggwkfgekrfseiasnterrtafvrsvapflrsygfdgldlawly 142
          ||  | :|||||::|||:||  :|::: : :: | :|| |   ||| |:|||||| | |
Qy     61 GLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEY 120

Db    143 p--r-lr--dkqyfstlikelnaeftkevq-pgreklllsaalsagkvaidtgydiaqia 196
          |  :     ||: |:||: :|   |  | | :|:|:|||||||::||   :|:||::  ||
Qy    121 PGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIA 180

Db    197 qhldfinlmtydfhgvwrqitghhsplfqgqkdtrfdrysnvnyavqymirlgaqaskll 256
          |:|||:|||:||||| |  :|||:|||:  | ::      ||: |||  :: |: ||||:
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    257 mgiptfgksftlass-enqlgapisgeglpgrftkeagtlayyeicdflkgaevhrlsne 315
          :|:||:|:|||||| : ::||| :| | || ||||:| |||||:|   |||  :|: ::
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQ 299

Db    316 kvpfatkgnqwvgyehkesvknkvgflkekklagamvwaldlddfqg-tcqpkeffpltn 374
          |||:  : |||||::  || | ||::||:| |:|||||||||||| | :|    : ||
Qy    300 KVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRY-PLIQ 358

Db    375 aikdala 381
          :::: |:
Qy    359 TLRQELS 365


RESULT    4
ENTRY             I48271     #type complete
TITLE             BRP39 protein - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
                    02-Jul-1996
ACCESSIONS        I48271
REFERENCE         I48271
   #authors       Morrison, B.W.; Leder, P.
   #journal       Oncogene (1994) 9:3417-3426
   #title         neu and ras initiate murine mammary tumors that share genetic
                    markers generally absent in c-myc and int-2-initiated
                    tumors.
   #cross-references MUID:95060797
   #accession     I48271
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-381 ##label RES
      ##cross-references EMBL:X93035; NID:g1085065; CDS_PID:g1085066
```

```
GENETICS
   #note          gene name brp39
SUMMARY           #length 381  #molecular-weight 43001  #checksum 5107

  Query Match           52.1%;  Score 1417;  DB 14;  Length 381;
  Best Local Similarity  50.7%;  Pred. No. 3.51e-237;
  Matches   186;  Conservative   84;  Mismatches  86;  Indels  11;  Gaps   9;

Db      23 klvcyftswsqyregvgsflpdaiqpflcthiiysfanissdnmlstwewndesnydkln 82
           |||||||:|:|||:| : |||  ::| ||||:||:|| :: :: ||| |||||: |: :|
Qy       2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMT-NHQLSTTEWNDETLYQEFN 60

Db      83 klktrntnlktllsvggwkfgekrfseiasnterrtafvrsvapflrsygfdgldlawly 142
            ||  | :|||||::|||:||  :|::: : :: | :|| |   ||| |:|||||| | |
Qy      61 GLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEY 120

Db     143 p--r-lr--dkqyfstlikelnaeftkevq-pgreklllsaalsagkvaidtgydiaqia 196
           |  :     ||: |:||: :|   |  | | :|:|:||||||::||   :|:||::  ||
Qy     121 PGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIA 180

Db     197 qhldfinlmtydfhgvwrqitghhsplfqggkdtrfdrysnvnyavqymirlgaqaskll 256
           |:|||:|||:||||| |  :|||:|||:  | ::      ||: |||  :: |: ||||:
Qy     181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db     257 mgiptfgksftlass-enqlgapisgeglpgrftkeagtlayyeicdflkgaevhrlsne 315
           :|:||:|:|||||| : ::||| :| | || ||||:| |||||:|   |||  :|: ::
Qy     241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQ 299

Db     316 kvpfatkgnqwvgyehkesvknkvgflkekklagamvwaldlddfqg-tcqpkeffpltn 374
           |||:  : |||||::  || | ||::||:| |:|||||||||||| | :|    : ||
Qy     300 KVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRY-PLIQ 358

Db     375 aikdala 381
           :::: |:
Qy     359 TLRQELS 365


RESULT    5
ENTRY             S27879     #type complete
TITLE             secretory protein YM-1 precursor - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
                    03-May-1996
ACCESSIONS        S27879
REFERENCE         S27879
   #authors       Chang, N.C.A.; Liu, C.H.; Chang, A.C.
   #submission    submitted to the EMBL Data Library, June 1992
   #description   Molecular characterization of a secretory protein (YM-1)
                    transiently expressed by activated murine peritoneal
                    macrophages.
   #accession     S27879
      ##molecule_type mRNA
      ##residues       1-399 ##label CHA
      ##cross-references EMBL:M94584
CLASSIFICATION    #superfamily Streptomyces chitinase chi40
FEATURE
   1-21                #domain signal sequence #status predicted #label SIG\
   22-399              #product secretory protein YM-1 #status predicted #label
                         MAT
SUMMARY           #length 399  #molecular-weight 44528  #checksum 3208

  Query Match           49.4%;  Score 1345;  DB 5;  Length 399;
  Best Local Similarity  47.7%;  Pred. No. 2.04e-223;
  Matches   175;  Conservative   87;  Mismatches 102;  Indels   3;  Gaps   3;

Db      24 lmcyytswakdrpiegsfkpgnidpclcthliyafagmqnneitytheqdlrdyealngl 83
           |:||:|:||  |  |: | | ::|| ||||||||||| |:::: |   |   |: :|||
Qy       3 LVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDETLYQEFNGL 62

Db      84 kdkkntelktllaiggwkfgpapfsamvstpqnrqifiqsvirflrqynfdglnldwqyp 143
           | |:|  |||||||||:||   |: ||:|: ||| |: | ||||| |:||||:|||:||
Qy      63 K-KMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db     144 gsrgsppkdkhlfsvlvkemrkafeeesvekdiprllltstgagiidviksgtrslncls 203
           ||:|||: ||  |: || :: :||::|:      ||||::: ::    : :|    :
Qy     122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQ 181

Db     204 lldyiqvmtydlhdpkdgytgensplykspydigksadlnvdsiisywkdhgaasekliv 263
           ||:: :|:||:| : :  || ||||||   : | :| ||||: :  | : |::: |||:
Qy     182 NLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLIL 241

Db     264 gfpayghtfilsdpsktgigaptistgppgkytdesgllayyevctflnegatevwdapq 323
           | |:||::| |: :| | :|||: ::| || :| |:|:|||||||:  : |||  :   |
Qy     242 GMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-K-GATKQRIQDQ 299

Db     324 evpyayqgnewvgydnvrsfklkaqwlkdnnlggavvwpldmddfsgsfchqrhfpltst 383
           ||| :: |:|||:|:| ||| |   ||:: |||||:||:||:|||:|  |:| ::||  |
Qy     300 KVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYPLIQT 359

Db     384 lkgdlni 390
           |: :|::
Qy     360 LRQELSL 366


RESULT    6
ENTRY             S57197     #type fragment
TITLE             oviduct-specific glycoprotein 95K precursor - bovine
                    (fragment)
ORGANISM          #formal_name Bos primigenius taurus #common_name cattle
DATE              28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
                    03-May-1996
ACCESSIONS        S57197
REFERENCE         S57197
   #authors       Sendai, Y.; Abe, H.; Kikuchi, M.; Satoh, T.; Hoshi, H.
   #journal       Biol. Reprod. (1994) 50:927-934
   #title         Purification and molecular cloning of bovine oviduct-specific
                    glycoprotein.
   #accession     S57197
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-537 ##label SEN
      ##cross-references EMBL:D16639
SUMMARY           #length 537  #checksum 1695

  Query Match           48.8%;  Score 1328;  DB 14;  Length 537;
  Best Local Similarity  51.2%;  Pred. No. 3.60e-220;
  Matches   187;  Conservative   74;  Mismatches  97;  Indels   7;  Gaps   6;

Db      20 klvcyftnwafsrpgpasilprdldpflcthlvfafasmsnnqivpkdpqdekilypefn 79
           ||||||||||  | | | :||:|||| |||||::|||:|:|:|:   :  ||  || |||
Qy       2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDET-LYQEFN 60

Db      80 klkernrglktllsiggwnfgtvrfttmlstfsnrerfvssviallrthgfdgldlffly 139
            ||  |  |||||:|||||||| :|| |::| :||: ||:| | :||  :||||||   |
Qy      61 GLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEY 120

Db     140 pglrgspardrwtfvflleellqafkneaqltmrprlllsaavsgdphvvqkayearllg 199
           || :|||| |:  |  |:::|  ||  ||| : : ||||||||::    |: :||   ::
Qy     121 PGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIA 180

Db     200 rlldfisvlsydlhgswekvtghnsplfs-l--pgdpks-sa-yamnywrqlgvppekll 254
           : |||:::::||:|||||||||||||||:     :| : | :   |:  | | | |: ||:
Qy     181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db     255 mglptygrtfhllkasqnelraqavgpaspgkytkqagflayyeiccfvrrakkrwindq 314
           :|:|||||:|  |  :|:  : | | |:::|| :||::| |||||:|   : | |::| ||
Qy     241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQ 299

Db     315 yvpyafkgkewvgyddaisfgykaffikrehfggamvwtldlddfrgyfcgtgpfplvht 374
           ||| |: ::|||:||  ||  |  ::|:  :||||||:|||||| |: |  | :||::|
Qy     300 KVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYPLIQT 359

Db     375 lnnll 379
           |   |
Qy     360 LRQEL 364
```

```
RESULT    7
ENTRY             I46470     #type complete
TITLE             estrogen dependent oviduct protein precursor - sheep
ORGANISM          #formal_name Ovis orientalis aries, Ovis ammon aries
                    #common_name domestic sheep
DATE              16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change
                    16-Aug-1996
ACCESSIONS        I46470
REFERENCE         I46470
   #authors       DeSouza, M.M.; Murray, M.K.
   #journal       Endocrinology (1995) 136:2485-2496
   #title         An estrogen-dependent secretory protein, which shares
                    identity with chitinases, is expressed in a temporally and
                    regionally specific manner in the sheep oviduct at the time
                    of fertilization and embryo development.
   #cross-references MUID:95269691
   #accession     I46470
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-539 ##label DES
      ##cross-references EMBL:U16719; NID:g885600; CDS_PID:g885601
SUMMARY           #length 539  #molecular-weight 59535  #checksum 2120

  Query Match            48.4%;  Score 1318;  DB 14;  Length 539;
  Best Local Similarity  50.1%;  Pred. No. 2.92e-218;
  Matches   183;  Conservative   77;  Mismatches  98;  Indels   7;  Gaps   5;

Db      23 klvcyftnwafsrpgsasilprdldpflcthlvfafasmnnnqivpkdpldekilypefn 82
           |||||||||||  | | | :||:|||| ||||||::|||:| |:|:   :  ::  || |||
Qy       2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTE-WNDETLYQEFN 60

Db      83 klkernrglktllsvggwnfgtsrftkmlstfsnrerfvksviallrthgfdgldlffly 142
             ||  |  |||||::|||||||| :|| |::| :||: ||:| | :||  :|||||   |
Qy      61 GLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEY 120

Db     143 pglrgspardrwtfvflleellqafkneaqltmrprlllsaavsgdphviqkaydarllg 202
           || :|||| |:  |  |:::|  ||  ||| : : |||||||::    :: :|:   ::
Qy     121 PGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIA 180

Db     203 rlldfisvlsydlhgswekvtghnsplfslpgdpkssa-yam--sy--wrqlgvppekll 257
           : |||:::::||:|||||||||||||||:    :: ::|   :  :   | | | |: ||:
Qy     181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db     258 mglptygrtfhllrasqnelgagaagpaspgkytkqagflayyevcsfvqrakkrwindq 317
           :|:|||||:| |  :|:  :|| |:|:::|| :||::| |||||||    | |::| ||
Qy     241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQ 299

Db     318 yvpyafkgkewvgyddaisfgykaffikrehfggamvwtldlddfrgnfcgtgpfplaht 377
           |||  |: ::|||:||  ||  |  ::|:  :||||||:|||||| |  |  | :|| :|
Qy     300 KVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYPLIQT 359

Db     378 lnnll 382
           |   |
Qy     360 LRQEL 364


RESULT    8
ENTRY             A53918     #type complete
TITLE             chitinase (EC 3.2.1.14) precursor - braconid wasp (Chelonus
                    sp.)
ORGANISM          #formal_name Chelonus sp.
DATE              28-Jul-1995 #sequence_revision 28-Jul-1995 #text_change
                    03-May-1996
ACCESSIONS        A53918
REFERENCE         A53918
   #authors       Krishnan, A.; Nair, P.N.; Jones, D.
   #journal       J. Biol. Chem. (1994) 269:20971-20976
   #title         Isolation, cloning, and characterization of new chitinase
                    stored in active form in chitin-lined venom reservoir.
   #accession     A53918
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-483 ##label KRI
      ##cross-references GB:U10422
KEYWORDS          glycosidase; hydrolase
SUMMARY           #length 483  #molecular-weight 52013  #checksum 5938

  Query Match            43.3%;  Score 1179;  DB 12;  Length 483;
  Best Local Similarity  42.5%;  Pred. No. 8.78e-192;
  Matches   158;  Conservative   98;  Mismatches  99;  Indels  17;  Gaps  11;

Db      23 kvvcyfgawsvyrqgngkfdingidptlcthliysfvgvngkdvkvldpwsdlpgnldgf 82
           |:||||  |: ||||:::|  : :||:|||||||:| |:   ::   : |:|    |  :
Qy       2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTE-WND--ETL--Y 56

Db      83 gkftslrkknpsvkimvavggwnagsvpfsqmasdqatreafaqnvvkflqqyqfdgfdi 142
             | :|:|:|| :| ::|:|||| |:  |::| :    |::|  : ::||: | |||:|:
Qy      57 QEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 116

Db     143 dweypaqrggspadvknmvklckalkkafvq--h----d-yilsaavaapetsasksydi 195
           |||||: :|::: |      |   | :|| |  :    :  :|||||:| :|    :|::
Qy     117 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 176

Db     196 aemsqyldfinlmtydfhgpwdghtgmhappsasshdsgnelklnvkaavkywlqngvpk 255
             ::| |||:|||:|||||:|:  || ::|      :||    ||| |||  |||:| |
Qy     177 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 236

Db     256 eklvvgvpaygksftlsnpsnkglgapvsgagtagpytgengllgyneicemqkagdwev 315
           ||::|:|:||:||||:::|:  :||| :|:|||:||:| | |:|:| |:|   |::  :
Qy     237 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQR 295

Db     316 vqdnekgvpyavkgnqwvsfddlaaikakaqfikqeglggamvwsietddfkgl-cge-k 373
           :|| :| |||  : ||||:|||: ::|:|  ::|| |||||||||::: ||| |: | : :
Qy     296 IQD-QK-VPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGR 353

Db     374 ypvlkalnsvlg 385
           ||:: :|   |:
Qy     354 YPLIQTLRQELS 365


RESULT    9
ENTRY             A38221     #type complete
TITLE             chitinase (EC 3.2.1.14) MF1 - nematode (Brugia malayi)
ORGANISM          #formal_name Brugia malayi
DATE              04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
                    03-May-1996
ACCESSIONS        A38221
REFERENCE         A38221
   #authors       Fuhrman, J.A.; Lane, W.S.; Smith, R.F.; Piessens, W.F.;
                    Perler, F.B.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1992) 89:1548-1552
   #title         Transmission-blocking antibodies recognize microfilarial
                    chitinase in brugian lymphatic filariasis.
   #cross-references MUID:92179220
   #accession     A38221
      ##status         preliminary; not compared with conceptual translation
      ##molecule_type nucleic acid; protein
      ##residues       1-504 ##label FUH
      ##cross-references NCBIP:85345
      ##note           sequence extracted from NCBI backbone
KEYWORDS          glycosidase; hydrolase
SUMMARY           #length 504  #molecular-weight 55971  #checksum 7343

  Query Match            39.9%;  Score 1086;  DB 12;  Length 504;
  Best Local Similarity  42.8%;  Pred. No. 3.90e-174;
  Matches   158;  Conservative   94;  Mismatches 102;  Indels  15;  Gaps  11;

Db      27 cyytnwaqyrdgegkflpgnipnglcthilyafakvdelgdskpfewndedtewskgmys 86
           ||:|||||||:||::||| ::  :||||::|||| : :   |   |||||  |     :|
Qy       5 CYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTT-EWNDE-T-----LYQ 57

Db      87 avtklretnpglkvllsyggynfgsaiftgiaksaqkterfiksaiaflrknnfdgfdld 146
             |:  || || ||: || |||:  || :  :| : : |::||| |||| :|||:|||
Qy      58 EFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 117
```

```
Db    147 weyp-v-gv-a-e-eh-aklveamktafveeaktsgkqrllltaavsagkgtidgsynve 200
          ||||   |  | : |: : ||: :  || :|| ||||:||||:|||:||   :|::|:|:
Qy    118 WEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVD 177

Db    201 slgknfdllflmsydlhgsweknvdlhgklhptkgevsgigifntefaadywaskgmpke 260
          :: |:|:: ||:||:|||||    :: |     | :: : :| : | : |  || |
Qy    178 KIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPAS 237

Db    261 kiiigipmyaqgwtldnpsetaigaaasrpssasktnpaggtasyweickylkeggketv 320
          |:|:|:| |::: || ::|:| :||:|: :::::     ||  :| |:|   | : |: :
Qy    238 KLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRI 296

Db    321 hqegvgaymvkgdqwygydneetirikmkwlkekgyggafiwaldfddftgkscgkgpyp 380
          ::: | :|: : :|| |:|: |::: |:  ||:|| ||| :||||:|||:| ||  | ||
Qy    297 QDQKV-PYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYP 355

Db    381 llnaissel 389
          |: ::  ||
Qy    356 LIQTLRQEL 364


RESULT   10
ENTRY            A56596     #type complete
TITLE            chitinase (EC 3.2.1.14) - tobacco hornworm
ORGANISM         #formal_name Manduca sexta #common_name tobacco hornworm
DATE             11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change
                   03-May-1996
ACCESSIONS       A56596
REFERENCE        A56596
   #authors      Kramer, K.J.; Corpuz, L.; Choi, H.K.; Muthukrishnan, S.
   #journal      Insect Biochem. Mol. Biol. (1993) 23:691-701
   #title        Sequence of a cDNA and expression of the gene encoding
                   epidermal and gut chitinases of Manduca sexta.
   #cross-references MUID:93357793
   #accession    A56596
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-554 ##label KRA
      ##cross-references NCBIN:136417; NCBIP:136418
      ##experimental_source larvae
      ##note          sequence extracted from NCBI backbone
KEYWORDS         glycosidase; hydrolase
SUMMARY          #length 554  #molecular-weight 62203  #checksum 4659

  Query Match           37.3%;  Score 1015;  DB 12;  Length 554;
  Best Local Similarity  40.4%;  Pred. No. 1.05e-160;
  Matches   155;  Conservative   98;  Mismatches 107;  Indels  24;  Gaps  18;

Db     23 arivcyfsnwavyrpgvgrygiedipvekcthiiysfigvtegnsevliidpeldvdkng 82
          |::||||:||| || | :|:   |:    |||:||:| |:|: : : |    | : |
Qy      1 AKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTN-H-Q-LST-TEWN-DETL 55

Db     83 frnftslrsshpsvkfmvavggwaegsskyshmvaqkstrmsfirsvvsflkkydfdgld 142
          :::| :|:  :| :| ::|:|||  |: |:: |||  : | :|: | : ||:|| |||||
Qy     56 YQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLD 115

Db    143 ldweypgaadrggsfsdkdkflylvqelrraf---irv-gkgwel-taavplanfrlmeg 197
          |||||||:  :|::  ||::|  |||:|  ||    :  || : | :|||| :   :  |
Qy    116 LDWEYPGS--QGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAG 173

Db    198 yhvpelcqeldaihvmsydlrgnwagfadvhsplykrphdqwayeklnvndglhlweekg 257
          | |  : |:|| :::|:||::|:|  :  :|||||  :  |   |||: ::: | :||
Qy    174 YEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKG 233

Db    258 cpsnklvvgipfygrsftlsagnnnyglgtfinkeagggdpapytnatgfwayyeictev 317
          |::||::|:| |||||||:::::   :|:     :|:| |:|:|:  |  ||||:|:
Qy    234 TPASKLILGMPTYGRSFTLASSSDT-RVGA--P-ATGSGTPGPFTKEGGMLAYYEVCS-- 287

Db    318 dkddsgwtkkwdeqgkcpyaykgtqwvgyedprsveikmnwikqkgylgamtwaidmddf 377
          |  :  || : :: | || ::  ||||::|  |   |:: :|||  ||| ||:|:|||
Qy    288 WKG-A--TKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDF 344

Db    378 qgl-cge-knplikilhkhms-sy 398
          |: | : : |||  |:  :| :|
Qy    345 AGFSCNQGRYPLIQTLRQELSLPY 368


RESULT   11
ENTRY            A38368     #type complete
TITLE            chitinase (EC 3.2.1.14) precursor - Bacillus circulans
ORGANISM         #formal_name Bacillus circulans
DATE             28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change
                   03-May-1996
ACCESSIONS       A38368
REFERENCE        A38368
   #authors      Watanabe, T.; Suzuki, K.; Oyanagi, W.; Ohnishi, K.; Tanaka,
                   H.
   #journal      J. Biol. Chem. (1990) 265:15659-15665
   #title        Gene cloning of chitinase A1 from Bacillus circulans WL-12
                   revealed its evolutionary relationship to Serratia
                   chitinase and to the type III homology units of
                   fibronectin.
   #cross-references MUID:90368776
   #accession    A38368
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-699 ##label WAT
      ##cross-references GB:J05599
CLASSIFICATION   #superfamily fibronectin type III repeat homology
KEYWORDS         glycosidase; hydrolase
SUMMARY          #length 699  #molecular-weight 73677  #checksum 6757

  Query Match           25.7%;  Score 698;  DB 10;  Length 699;
  Best Local Similarity  40.1%;  Pred. No. 2.18e-101;
  Matches   119;  Conservative   73;  Mismatches 85;  Indels  20;  Gaps  18;

Db    144 qlnklkqtnpnlktiisvggwtw-snrfsdvaataatrevfansavdflrkynfdgvdld 202
          ::| ||  ||:|||::::|||   : :|:|: |||  |: | |||: |||||:|||:|||
Qy     58 EFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 117

Db    203 weypvsggldgnskrpedkqnytlllskirekldaagavdgk-kylltiasgasatya-a 260
          ||||  |: :| |  : ||: :| |:  : : ::  :   || : ||: |  |: ||  |
Qy    118 WEYP--GS-QG-S-PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDA 172

Db    261 ntelakiaaivdwinimtydfngawqkisahnaplnydpaasaagvpdantfnvaagaqg 320
            |: |||  :| :|:|:|||:|:|:|:::||:|| |      :| : | ::|| |: |
Qy    173 GYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPL-YKRQE-ESGAA-A-SLNVDAAVQQ 228

Db    321 hldagvpaaklvlgvpfygrgwd-gcaqagn-gqyqtctggssvgtweagsfdfydlean 378
          |: | ||:||:||:| |||:   : :     |   | :| ::  | |:| : :|::  :
Qy    229 WLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEV-CS 287

Db    379 yinkngytrywndtakvpylynasnkrfisyddaesvgyktayikskglggamfwel 435
            | |: : :  ||||::   | : :::|| ||   | :|:| |||||| | |
Qy    288 W--K-GATKQRIQDQKVPYIFRD-NQW-VGFDDVESFKTKVSYLKQKGLGGAMVWAL 339


RESULT   12
ENTRY            S51591     #type complete
TITLE            chitinase (EC 3.2.1.14)/lysozyme (EC 3.2.1.17) PZ precursor,
                   pathogenesis-related - common tobacco
ORGANISM         #formal_name Nicotiana tabacum #common_name common tobacco
DATE             15-Jul-1995 #sequence_revision 08-Sep-1995 #text_change
                   03-May-1996
ACCESSIONS       S51591; S51632; S43119
REFERENCE        S51591
   #authors      Heitz, T.; Segond, S.; Kauffmann, S.; Geoffroy, P.; Prasad,
                   V.; Brunner, F.; Fritig, B.; Legrand, M.
   #journal      Mol. Gen. Genet. (1994) 245:246-254
   #title        Molecular characterization of a novel tobacco
                   pathogenesis-related (PR) protein: a new plant
                   chitinase/lysozyme.
   #accession    S51591
      ##molecule_type mRNA
      ##residues      1-378 ##label HEI
      ##cross-references EMBL:X78325
```

```
     ##experimental_source cv. Samsun NN
   #accession    S51632
     ##molecule_type protein
     ##residues      31-36;87-112;252-275;282-305;337-371 ##label HEW
CLASSIFICATION   #superfamily Streptomyces chitinase chi40
KEYWORDS         glycosidase; hydrolase
FEATURE
   1-25               #domain signal sequence #status predicted #label SIG\
   26-371             #product chitinase/lysozyme PZ #status predicted #label
                       MAT\
   371-378            #domain carboxyl-terminal propeptide #status
                       experimental #label PRO
SUMMARY          #length 378  #molecular-weight 42019  #checksum 6689

  Query Match           21.8%;  Score 594;  DB 5;  Length 378;
  Best Local Similarity  32.7%;  Pred. No. 2.86e-82;
  Matches   106;  Conservative   87;  Mismatches 108;  Indels  23;  Gaps  18;

Db     43 nnidstlfthlfcafadlnpqsnqliispenqds-fsqftstvqrknpsvktflsiaggr 101
          :::|::| |||: ||| :   ::||  :  |::: : :| : : ::|| :||:|:| ||:
Qy     23 KDLDPSLCTHLIYAFAGMT--NHQLSTTEWNDETLYQEFNG-LKKMNPKLKTLLAI-GGW 78

Db    102 a-dttaygimarqpnsrksfidssirlarqfgfhgldldweyp--l-s-atdmtnlgill 156
            |  :  |   :|:| :|::|:||: | ::| ||||||||    | | |:  :  |:
Qy     79 NFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLV 138

Db    157 newrtainmearnsgraallltaavsysprvnglnypvesvarnlnwinlmaydfygpnw 216
           :   |:  ||: ||:  |||:|||: :       | |: :|:||: :|||||| |: |
Qy    139 QDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGS-W 197

Db    217 spsqtnshaqlfdpvnh--isgs---dg-inawiqagvptkklvlgipfygyawrlvnpn 270
             |  :: |:   :    ::|   |: :  |:| | |: ||:||:| || :  | :::
Qy    198 EKV-TGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSS 256

Db    271 ihdlrapaagksnvga-vddgsm-tynrirdyivqsrattvynativgdycysgsnwisy 328
           : |||:| :  |:   :|:| :|  :       ::   : :   |  | :  : |:::
Qy    257 DTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQR-IQDQK-VP-YIFRDNQWVGF 313

Db    329 ddtqsvrnkvnyvkgrgllgyfaw 352
          || :| : ||:|:| :|| |   |
Qy    314 DDVESFKTKVSYLKQKGLGGAMVW 337


RESULT   13
ENTRY            S47133     #type complete
TITLE            chitinase (EC 3.2.1.14) - fungus (Trichoderma harzianum)
ORGANISM         #formal_name Trichoderma harzianum
DATE             23-Nov-1994 #sequence_revision 03-Aug-1995 #text_change
                  10-May-1996
ACCESSIONS       S47133
REFERENCE        S47133
   #authors      Carsolio, C.; Gutierrez, A.; Jimenez, B.; van Montagu, M.;
                  Herrera-Estrella, A.
   #submission   submitted to the EMBL Data Library, June 1994
   #description  Characterization of ech-42, a Trichoderma harzianum
                  endochitinase gene expressed during mycoparasitism.
   #accession    S47133
     ##status         preliminary
     ##molecule_type DNA
     ##residues      1-424 ##label CAR
     ##cross-references EMBL:X79381
GENETICS
   #introns      48/2; 81/2; 98/1
CLASSIFICATION   #superfamily Streptomyces chitinase chi40
KEYWORDS         glycosidase; hydrolase
SUMMARY          #length 424  #molecular-weight 46298  #checksum 8773

  Query Match           21.4%;  Score 582;  DB 5;  Length 424;
  Best Local Similarity  36.0%;  Pred. No. 4.43e-80;
  Matches   102;  Conservative   69;  Mismatches  95;  Indels  17;  Gaps  13;

Db    112 qlfk-lkkanrnlkvmlsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidv 169
          | |: ||| | :|| :|:|||| :| :|    :|  ||  | ::|| |::  :|||:|:
Qy     57 QEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 116

Db    170 dweyp-addtqatnmvlllkeirsqldayaaqy-apg-yhfllsiaapagpehy-sflhm 225
          ||||| :: : | ::  :   ::   :|:  :  ::|  ::||| | |||     :   :
Qy    117 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 176

Db    226 sdlgqvldyvnlmaydyagswssysghdanlfanpsnpnssp-yntdqaikdyikggvpa 284
            ::| ||:|||||||: |||   :||:: |:    :: ::  | | |: : :  | ||
Qy    177 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 236

Db    285 skivlgmpiygraf--estgg--igqtysgigs-gsw--engiwdy-kvlpkagatvqyd 336
          ||::|||| |||:|   |::  :|  :| |: |:   | |:  |  | :  ||| |
Qy    237 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQ-- 294

Db    337 svaqayysydpsskelisfdtpdmintkvsylknlglggsmfw 379
           : :   :|    :: ::||  : ::|||||||  ||||:| |
Qy    295 RIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT   14
ENTRY            JQ1975     #type complete
TITLE            chitinase (EC 3.2.1.14) 1 precursor - imperfect fungus
                  (Aphanocladium album)
ORGANISM         #formal_name Aphanocladium album
DATE             03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change
                  03-May-1996
ACCESSIONS       JQ1975
REFERENCE        JQ1975
   #authors      Blaiseau, P.L.; Lafay, J.F.
   #journal      Gene (1992) 120:243-248
   #title        Primary structure of a chitinase-encoding gene (chi1) from
                  the filamentous fungus Aphanocladium album: Similarity to
                  bacterial chitinases.
   #accession    JQ1975
     ##molecule_type mRNA
     ##residues      1-423 ##label BLA
     ##cross-references GB:X64104
COMMENT     This enzyme is essential for the degradation of insoluble chitin.
GENETICS
   #gene         chi1
   #introns      47/2; 80/2; 97/1
CLASSIFICATION   #superfamily Streptomyces chitinase chi40
KEYWORDS         glycosidase; hydrolase
FEATURE
   1-34               #domain signal sequence #status predicted #label SIG\
   35-423             #product chitinase 1 #status predicted #label MAT
SUMMARY          #length 423  #molecular-weight 46072  #checksum 8802

  Query Match           21.2%;  Score 576;  DB 5;  Length 423;
  Best Local Similarity  36.2%;  Pred. No. 5.50e-79;
  Matches   101;  Conservative   65;  Mismatches  95;  Indels  18;  Gaps  16;

Db    115 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 173
          ||| | ::| :|:|||| :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    174 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 228
           :::: | ::  :   |:   :::  :  : |  ::||| | ||| : |    :  : ::
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 180

Db    229 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 287
            ||::||||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    288 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 339
          |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 300

Db    340 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 378
            | :    :: ::||    : |||| || |||||:| |
Qy    301 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337
```

```
RESULT   15
ENTRY            S51369     #type complete
TITLE            chitinase - fungus (Trichoderma harzianum)
ORGANISM         #formal_name Trichoderma harzianum
DATE             15-Jul-1995 #sequence_revision 19-Apr-1996 #text_change
                   03-May-1996
ACCESSIONS       S51369
REFERENCE        S51369
   #authors      Garcia, I.; Lora, J.M.; de la Cruz, J.; Benitez, T.; Llobell,
                   A.; Pintor-Toro, J.A.
   #journal      Curr. Genet. (1994) 27:83-89
   #title        Cloning and characterization of a chitinase (CHIT42) cDNA
                   from the mycoparasitic fungus Trichoderma harzianum.
   #accession    S51369
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-423 ##label GAR
SUMMARY          #length 423  #molecular-weight 46056  #checksum 2000

  Query Match           20.6%;  Score 561;  DB 12;  Length 423;
  Best Local Similarity  35.7%;  Pred. No. 2.95e-76;
  Matches   101;  Conservative   71;  Mismatches  94;  Indels  17;  Gaps  13;

Db    111 qlfk-vkkanrglkvllsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidi 168
          | |: :|| |  || ||:|||| :| :|    :|  ||  | ::|| |::  :|||:|:
Qy     57 QEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 116

Db    169 dweyp-adatqasnmilllkevrsqrdayaaqy-apg-yhflltiaapagkdny-sklrl 224
          ||||| :::: | ::  :   |:   :|:  :  ::|  ::||: | |||     :   :
Qy    117 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 176

Db    225 adlgqvldyinlmaydyagsfspltghdanlfnnpsnpnatp-fntdsavkdyinggvpa 283
            ::| ||::|||||: ||   :|||:: |::   :: |:: :| |:|| : :  | ||
Qy    177 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 236

Db    284 nkivlgmpiygrsf--qntag--igqtyngvgs-gsw--eagiwdykalpk-agatvqyd 335
          :|::|||| |||||   :::   :|    | |: |:   |:|:  |  :    ||| |
Qy    237 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRI 296

Db    336 svakgyysynsatkelisfdtpdmintkvaylkslglggsmfw 378
            |  | :    :: ::||  : ::|||:|||  ||||:| |
Qy    297 QDQKVPYIFRD--NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


Search completed: Fri Jun 27 17:02:53 1997
Job time : 113 secs.
```

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 27 16:59:40 1997; MasPar time 12.05 Seconds
656.435 Million cell updates/sec
Tabular output not generated.

Title: >US-08-663-618A-14
Description: (1-373) from US08663618A.pep
Perfect Score: 2721
Sequence: 1 AKLVCYFTNWAQYRQGEARF.....YPLIQTLRQELSLPYLPSGT 373

Scoring table: PAM 150
Gap 11

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot34
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11

Statistics: Mean 48.978; Variance 85.387; scale 0.574

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1488	54.7	383	4	GP39_HUMAN	CARTILAGE GLYCOPROTEI	0.00e+00
2	1086	39.9	504	2	CHIT_BRUMA	ENDOCHITINASE PRECURS	6.23e-219
3	1015	37.3	554	2	CHIT_MANSE	ENDOCHITINASE PRECURS	4.86e-202
4	698	25.7	699	2	CHI1_BACCI	CHITINASE A1 PRECURSO	1.82e-127
5	576	21.2	423	2	CHI1_APHAL	CHITINASE 1 PRECURSOR	2.50e-99
6	561	20.6	423	2	CHI4_TRIHA	42 KD ENDOCHITINASE P	6.68e-96
7	546	20.1	427	2	CHI1_COCIM	ENDOCHITINASE 1 PRECU	1.75e-92
8	437	16.1	563	2	CHIA_SERMA	CHITINASE A PRECURSOR	6.12e-68
9	395	14.5	551	2	CHIT_NPVAC	PROBABLE ENDOCHITINAS	1.16e-58
10	345	12.7	820	2	CHIA_ALTSO	CHITINASE A PRECURSOR	8.67e-48
11	337	12.4	499	2	CHIB_SERMA	CHITINASE B PRECURSOR	4.53e-46
12	301	11.1	619	2	CHIT_STRLI	CHITINASE C PRECURSOR	2.01e-38
13	287	10.5	610	2	CHIT_STRPL	CHITINASE 63 PRECURSO	1.72e-35
14	263	9.7	1146	5	KTXA_KLULA	KILLER TOXIN ALPHA AN	1.56e-30
15	172	6.3	36	4	GP39_BOVIN	CARTILAGE GLYCOPROTEI	8.27e-13
16	172	6.3	83	3	EDOS_PAPAN	ESTRADIOL-DEPENDENT O	8.27e-13
17	121	4.4	597	2	CHIX_STROI	EXOCHITINASE 1 PRECUR	3.59e-04
18	108	4.0	452	4	HEMF_RHOSH	COPROPORPHYRINOGEN II	3.08e-02
19	110	4.0	497	2	CHID_BACCI	CHITINASE D PRECURSOR	1.59e-02
20	104	3.8	422	10	UMUC_ECOLI	UMUC PROTEIN.	1.13e-01
21	100	3.7	364	11	YBR5_YEAST	HYPOTHETICAL 40.9 KD	3.99e-01
22	101	3.7	1045	7	PRTT_SERMA	EXTRACELLULAR SERINE	2.92e-01
23	101	3.7	1070	1	AGLU_CANTS	ALPHA-GLUCOSIDASE PRE	2.92e-01
24	99	3.6	398	9	SPEB_STRPY	EXOTOXIN TYPE B PRECU	5.44e-01
25	98	3.6	492	9	SYK_THETH	LYSYL-TRNA SYNTHETASE	7.40e-01
26	97	3.6	563	5	K2CB_HUMAN	KERATIN, TYPE II CYTO	1.00e+00
27	97	3.6	563	5	K2CA_HUMAN	KERATIN, TYPE II CYTO	1.00e+00
28	97	3.6	563	5	K2CC_HUMAN	KERATIN, TYPE II CYTO	1.00e+00
29	99	3.6	958	1	AMYG_SCHOC	GLUCOAMYLASE 1 PRECUR	5.44e-01
30	98	3.6	1045	7	PRTS_SERMA	EXTRACELLULAR SERINE	7.40e-01
31	96	3.5	297	5	HMX1_HUMAN	HOMEOBOX PROTEIN MSX-	1.36e+00
32	94	3.5	344	4	GCH2_PICGU	GTP CYCLOHYDROLASE II	2.46e+00
33	95	3.5	353	11	YQHT_BACSU	PUTATIVE PEPTIDASE IN	1.83e+00
34	94	3.5	404	6	MOEA_HAEIN	MOLYBDOPTERIN BIOSYNT	2.46e+00
35	95	3.5	492	7	PDI_SCHPO	PUTATIVE PROTEIN DISU	1.83e+00
36	96	3.5	554	5	HYES_RAT	SOLUBLE EPOXIDE HYDRO	1.36e+00
37	95	3.5	554	5	HYES_MOUSE	SOLUBLE EPOXIDE HYDRO	1.83e+00
38	95	3.5	563	5	K2CE_HUMAN	KERATIN, TYPE II CYTO	1.83e+00
39	95	3.5	563	5	K2CF_HUMAN	KERATIN, TYPE II CYTO	1.83e+00
40	94	3.5	993	11	YAJ1_SCHPO	PUTATIVE FAMILY 31 GL	2.46e+00
41	95	3.5	1333	10	VGR1_MOUSE	VASCULAR ENDOTHELIAL	1.83e+00
42	93	3.4	267	3	EBAG_FLASP	ENDO-BETA-N-ACETYLGLU	3.30e+00
43	93	3.4	295	11	YHO8_YEAST	HYPOTHETICAL 35.1 KD	3.30e+00
44	92	3.4	510	1	BMP6_MOUSE	BONE MORPHOGENETIC PR	4.41e+00
45	92	3.4	557	3	ERK3_HUMAN	EXTRACELLULAR SIGNAL-	4.41e+00

ALIGNMENTS

RESULT 1
ID GP39_HUMAN STANDARD; PRT; 383 AA.
AC P36222; P30923;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE CARTILAGE GLYCOPROTEIN-39 PRECURSOR (GP-39) (39 KD SYNOVIAL PROTEIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=CARTILAGE;
RX MEDLINE; 94064658.
RA HAKALA B.E., WHITE C., RECKLIES A.D.;
RL J. BIOL. CHEM. 268:25803-25810(1993).
RN [2]
RP SEQUENCE OF 22-45.
RX MEDLINE; 90328983.
RA NYIRKOS P., GOLDS E.E.;
RL BIOCHEM. J. 269:265-268(1990).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE CAPACITY OF CELLS TO
CC RESPOND TO AND COPE WITH CHANGES IN THEIR ENVIRONMENT.
CC -!- TISSUE SPECIFICITY: PRESENT IN ARTICULAR CHONDROCYTES, SYNOVIAL
CC CELLS AS WELL AS IN LIVER. UNDETECTABLE IN MUSCLE TISSUES, LUNG,
CC PANCREAS, MONONUCLEAR CELLS, OR FIBROBLASTS.
CC -!- PTM: GLYCOSYLATED.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
DR EMBL; M80927; G348912; -.
DR PIR; S10677; S10677.
DR PIR; A33162; A33162.
DR PROSITE; PS01095; CHITINASE_18.
KW GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 21
FT CHAIN 22 383 CARTILAGE GLYCOPROTEIN-39.
FT CARBOHYD 60 60 POTENTIAL.
SQ SEQUENCE 383 AA; 42613 MW; D0A5280F CRC32;

Query Match 54.7%; Score 1488; DB 4; Length 383;
Best Local Similarity 53.3%; Pred. No. 0.00e+00;
Matches 195; Conservative 83; Mismatches 79; Indels 9; Gaps 5;

```
Db      23 klvcyytswsqyregdgscfpdaldrflcthiiysfanisndhidtwewndvtlygmlnt 82
           |||||:|:|:|||:|::  :|  ||  ||||:||:|| ::| :: | |||| |||  :|
Qy       2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDETLYQEFNG 61
```

```
Db     83 lknrnpnlktllsvggwnfgsqrfskiasntqsrrtfiksvppflrthgfdgldlawlyp 142
          ||: ||:|||||::||||||:|:|: : : : :|:||::|   |||  :|||||| | ||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    143 grr-----dkqhfttlikemkaefikeaq-pgkkqlllsaalsagkvtidssydiakisq 196
          | :     ||::||||: ::   |  ||| :|| :||||||::||   :|::|:: ||:|
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQ 181

Db    197 hldfisimtydfhgawrgttghhsplfrgqedaspdrfsntdyavgymlrlgapasklvm 256
          :|||:::|:|||||:|   |||:|||:: ||:::    | | ||   |: |:|||||::
Qy    182 NLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLIL 241

Db    257 giptfgrsftlass-etgvgapisgpgipgrftkeagtlayyeicdflrgatvhrtlgqq 315
          |:||:|||||||| :| |||| :|:| || ||||:| |||||:|   :||| :|   |
Qy    242 GMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQK 300

Db    316 vpyatkgnqwvgyddqesvkskvqylkdrqlagamvwaldlddfqgsfcgqdlrfpltna 375
          |||  : |||||:|| || |:|| |||:: |:|||||||||||| |  | |  |:||  :
Qy    301 VPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQG-RYPLIQT 359

Db    376 ikdala 381
          ::: |:
Qy    360 LRQELS 365


RESULT    2
ID   CHIT_BRUMA     STANDARD;      PRT;   504 AA.
AC   P29030;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   ENDOCHITINASE PRECURSOR (EC 3.2.1.14) (MF1 ANTIGEN).
OS   BRUGIA MALAYI.
OC   EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; SPIRURIDA.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE; 92179220.
RA   FUHRMAN J.A., LANE W.S., SMITH R.F., PIESSENS W.F., PERLER F.B.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 89:1548-1552(1992).
CC   -!- FUNCTION: THE MF1 ANTIGEN IS A MICROFILARIAL CHITINASE, WHICH MAY
CC       FUNCTION TO DEGRADE CHITIN-CONTAINING STRUCTURES IN THE MICRO-
CC       FILARIA OR IN ITS MOSQUITO VECTOR DURING PARASITE DEVELOPMENT AND
CC       TRANSMISSION.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- PTM: O-GLYCOSYLATED.
CC   -!- KNOWN TO BIND CALCIUM.
CC   -!- DEVELOPMENTAL STAGE: THE APPEARANCE OF THE MF1 ANTIGEN CORRESPOND
CC       WITH THE ONSET OF THE PARASITE'S ABILITY TO INFECT THE MOSQUITO.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; M73689; G156064; -.
DR   PIR; A38221; A38221.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ANTIGEN;
KW   REPEAT; GLYCOPROTEIN; CALCIUM-BINDING.
FT   SIGNAL        1     22
FT   CHAIN        23    504       ENDOCHITINASE.
FT   DOMAIN       23    400       CATALYTIC.
FT   DOMAIN      401    450       SER/THR-RICH (LINKER).
FT   DOMAIN      407    448       3 X 14 AA APPROXIMATE TANDEM REPEATS.
FT   ACT_SITE    148    148       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   504 AA;  55971 MW;  4DA7E5EB CRC32;

  Query Match           39.9%;  Score 1086;  DB 2;  Length 504;
  Best Local Similarity  42.8%;  Pred. No. 6.23e-219;
  Matches   158;  Conservative   94;  Mismatches 102;  Indels  15;  Gaps  11;

Db     27 cyytnwaqyrdgegkflpgnipnglcthilyafakvdelgdskpfewndedtewskgmys 86
          ||:|||||||:||::||| ::  :||||::|||| : :   |   ||||| |     :|
Qy      5 CYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTT-EWNDE-T-----LYQ 57

Db     87 avtklretnpglkvllsyggynfgsaiftgiaksaqkterfiksaiaflrknnfdgfdld 146
          |:  || || ||: || |||:  || :  :| : : |::||| |||| :|||:|||
Qy     58 EFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 117

Db    147 weyp-v-gv-a-e-eh-aklveamktafveeaktsgkqrllltaavsagkgtidgsynve 200
          ||||   |  | : |: : ||: :  || :|| ||||:||||:|||:||   :|::|:|:
Qy    118 WEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVD 177

Db    201 slgknfdllflmsydlhgsweknvdlhgklhptkgevsgigifntefaadywaskgmpke 260
          :: |:|:: ||:||:|||||    :: |     | :: : :| : | : |  || |
Qy    178 KIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPAS 237

Db    261 kiiigipmyaqgwtldnpsetaigaaasrpssasktnpaggtasyweickylkeggketv 320
          |:|:|:| |::: || ::|:| :||:|: ::::     ||  :| |:|   | : |: :
Qy    238 KLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRI 296

Db    321 hqegvgaymvkgdqwygydneetirikmkwlkekgyggafiwaldfddftgkscgkgpyp 380
          ::: | :|: : :|| |:|: |::: |:  ||:|| ||| :||||:|||:| ||  | ||
Qy    297 QDQKV-PYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYP 355

Db    381 llnaissel 389
          |: ::  ||
Qy    356 LIQTLRQEL 364


RESULT    3
ID   CHIT_MANSE     STANDARD;      PRT;   554 AA.
AC   P36362;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
OS   MANDUCA SEXTA (TOBACCO HAWKMOTH) (TOBACCO HORNWORM).
OC   EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; LEPIDOPTERA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93357793.
RA   KRAMER K.J., CORPUZ L., CHOI H.K., MUTHUKRISHNAN S.;
RL   INSECT BIOCHEM. MOL. BIOL. 23:691-701(1993).
CC   -!- FUNCTION: DIGEST CHITIN IN THE EXOSKELETON DURING THE MOLTING
CC       PROCESS.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- DEVELOPMENTAL STAGE: HIGH LEVELS SEEN IN THE EPIDERMIS ON DAY 0,
CC       BUT RAPIDLY DISAPPEARS AND IS UNDETECTED ON DAYS 1-4 OF FIFTH
CC       INSTAR. IT REAPPEARS ON DAY 5 AND PEAKS ON DAY 7 AFTER WHICH A
CC       RAPID DECLINE IS SEEN. IN THE GUT IS DETECTED ON DAY 6 WITH LOWER
CC       LEVELS SEEN ON DAYS 0, 7 AND 8.
CC   -!- TISSUE SPECIFICITY: EPIDERMIS AND GUT.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; U02270; G406049; -.
DR   EMBL; S64757; -; NOT_ANNOTATED_CDS.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    554       ENDOCHITINASE.
FT   DOMAIN      396    453       SER/THR-RICH.
FT   ACT_SITE    146    146       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD     85     85       POTENTIAL.
FT   CARBOHYD    303    303       POTENTIAL.
FT   CARBOHYD    407    407       POTENTIAL.
FT   CARBOHYD    545    545       POTENTIAL.
SQ   SEQUENCE   554 AA;  62203 MW;  FA87F8AD CRC32;

  Query Match           37.3%;  Score 1015;  DB 2;  Length 554;
  Best Local Similarity  40.4%;  Pred. No. 4.86e-202;
  Matches   155;  Conservative   98;  Mismatches 107;  Indels  24;  Gaps  18;

Db     23 arivcyfsnwavyrpgvgrygiedipvekcthiiysfigvtegnsevliidpeldvdkng 82
          |::||||:||| || | :|:   |:    |||:||:| |:|: : : |    | : |
```

```
Qy       1 AKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTN-H-Q-LST-TEWN-DETL 55

Db      83 frnftslrsshpsvkfmvavggwaegsskyshmvaqkstrmsfirsvvsflkkydfdgld 142
           :::| :|:  :| :| ::|:|||  |: |:: ||| : | :|: |.: ||:|| |||||
Qy      56 YQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLD 115

Db     143 ldweypgaadrggsfsdkdkflylvqelrraf---irv-gkgwel-taavplanfrlmeg 197
           |||||||:  :|::  ||::|  |||:|  ||    :  || : | :|||| :   :  |
Qy     116 LDWEYPGS--QGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAG 173

Db     198 yhvpelcqeldaihvmsydlrgnwagfadvhsplykrphdqwayeklnvndglhlweekg 257
           | |  : |:|| :::|:|||::|:|   :  :|||||  :  |   |||: ::: | :||
Qy     174 YEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKG 233

Db     258 cpsnklvvgipfygrsftlsagnnnyglgtfinkeagggdpapytnatgfwayyeictev 317
           |::||::|:| |||||||:::::   :|:     :|:| |:|:|:  |  ||||:|:
Qy     234 TPASKLILGMPTYGRSFTLASSSDT-RVGA--P-ATGSGTPGPFTKEGGMLAYYEVCS-- 287

Db     318 dkddsgwtkkwdeqgkcpyaykgtqwvgyedprsveikmnwikqkgylgamtwaidmddf 377
           |  :  || : :: | || ::  ||||::|  |   |:: :|||| ||| ||:|:|||
Qy     288 WKG-A--TKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDF 344

Db     378 qgl-cge-knplikilhkhms-sy 398
           |: | : : |||  |:  :| :|
Qy     345 AGFSCNQGRYPLIQTLRQELSLPY 368


RESULT    4
ID   CHI1_BACCI     STANDARD;      PRT;   699 AA.
AC   P20533;
DT   01-FEB-1991 (REL. 17, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE A1 PRECURSOR (EC 3.2.1.14).
GN   CHIA1.
OS   BACILLUS CIRCULANS.
OC   PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WL-12;
RX   MEDLINE; 90368776.
RA   WATANABE T., SUZUKI K., OYANAGI W., OHNISHI K., TANAKA H.;
RL   J. BIOL. CHEM. 265:15659-15665(1990).
RN   [2]
RP   MUTAGENESIS.
RC   STRAIN=WL-12;
RX   MEDLINE; 93366760.
RA   WATANABE T., KOHORI K., MIYASHITA K., FUJII T., SAKAI H.,
RA   UCHIDA M., TANAKA H.;
RL   J. BIOL. CHEM. 268:18567-18572(1993).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; M57601; G142688; -.
DR   PIR; A38368; A38368.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; REPEAT.
FT   SIGNAL        1     41
FT   CHAIN        42    699       CHITINASE A1.
FT   DOMAIN       42    460       CATALYTIC.
FT   DOMAIN      465    549       FIBRONECTIN TYPE-III (R-1).
FT   DOMAIN      560    644       FIBRONECTIN TYPE-III (R-2).
FT   ACT_SITE    204    204       PROTON DONOR (PROBABLE).
FT   MUTAGEN     200    200       D->N: DECREASE IN ACTIVITY.
FT   MUTAGEN     200    200       D->E: NO CHANGE IN ACTIVITY.
FT   MUTAGEN     204    204       E->D,Q: LOSS OF ACTIVITY.
SQ   SEQUENCE   699 AA;  73677 MW;  5A1777CC CRC32;

  Query Match           25.7%;  Score 698;  DB 2;  Length 699;
  Best Local Similarity 40.1%;  Pred. No. 1.82e-127;
  Matches  119;  Conservative   73;  Mismatches  85;  Indels  20;  Gaps  18;

Db     144 qlnklkqtnpnlktiisvggwtw-snrfsdvaataatrevfansavdflrkynfdgvdld 202
           ::| ||  ||:|||:::|||   : :|:|: |||  |: | |||: |||||:|||:|||
Qy      58 EFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 117

Db     203 weypvsggldgnskrpedkqnytlllskirekldaagavdgk-kylltiasgasatya-a 260
           ||||  |: :| |  : ||: :| |:  : : ::  :   || : ||: |  |: ||  |
Qy     118 WEYP--GS-QG-S-PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDA 172

Db     261 ntelakiaaivdwinimtydfngawqkisahnaplnydpaasaagvpdantfnvaagaqg 320
           |: |||  :| :|:|:|||:|:|:|:::||:|| |      :| : | ::|| |: |
Qy     173 GYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPL-YKRQE-ESGAA-A-SLNVDAAVQQ 228

Db     321 hldagvpaaklvlgvpfygrgwd-gcaqagn-gqyqtctggssvgtweagsfdfydlean 378
           |: | ||:||:||:| |||:   : :      |   | :| ::  | |:| : :|::  :
Qy     229 WLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEV-CS 287

Db     379 yinkngytrywndtakvpylynasnkrfisyddaesvgyktayikskglggamfwel 435
           | | |: : :  ||||::   | : :::|| ||   | :|:| |||||||  | |
Qy     288 W--K-GATKQRIQDQKVPYIFRD-NQW-VGFDDVESFKTKVSYLKQKGLGGAMVWAL 339


RESULT    5
ID   CHI1_APHAL     STANDARD;      PRT;   423 AA.
AC   P32470;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   CHITINASE 1 PRECURSOR (EC 3.2.1.14).
GN   CHI1.
OS   APHANOCLADIUM ALBUM.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ETHM 483;
RX   MEDLINE; 93013040.
RA   BLAISEAU P.-L., LAFAY J.-F.;
RL   GENE 120:243-248(1992).
RN   [2]
RP   SEQUENCE OF 35-57.
RX   MEDLINE; 92136437.
RA   BLAISEAU P.-L., KUNZ C., GRISON R., BERTHEAU Y., BRYGOO Y.;
RL   CURR. GENET. 21:61-66(1992).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; X64104; G429026; -.
DR   PIR; JQ1975; JQ1975.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ZYMOGEN.
FT   SIGNAL        1     22       POTENTIAL.
FT   PROPEP       23     34       POTENTIAL.
FT   CHAIN        35    423       CHITINASE 1.
FT   ACT_SITE    171    171       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   423 AA;  46058 MW;  88456E6E CRC32;

  Query Match           21.2%;  Score 576;  DB 2;  Length 423;
  Best Local Similarity 36.2%;  Pred. No. 2.50e-99;
  Matches  101;  Conservative   65;  Mismatches  95;  Indels  18;  Gaps  16;

Db     115 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 173
           ||| | ::| :|:|||| :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy      62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db     174 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 228
            :::: | ::  :  |:  ::: :  : |  ::||| | ||| : |    :  : ::
Qy     122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 180

Db     229 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 287
```

```
           ||::||||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    288 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 339
          |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 300

Db    340 gcysydpstkelisfdtpamistkvswlkgkglggtmfw 378
            | :    :: ::||    : |||| || |||||:| |
Qy    301 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT    6
ID   CHI4_TRIHA     STANDARD;      PRT;   423 AA.
AC   P48827;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   42 KD ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
GN   CHIT42.
OS   TRICHODERMA HARZIANUM.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 35-52; 93-107; 371-385 & 397-414.
RX   MEDLINE; 95269313.
RA   GARCIA I., LORA J.M., LA CRUZ J., BENITEZ T., LLOBELL A.,
RA   PINTOR-TORO J.A.;
RL   CURR. GENET. 27:83-89(1994).
CC   -!- FUNCTION: MORPHOGENETIC ROLE DURING APICAL GROWTH,CELL DIVISION
CC       AND DIFFERENTIATION (CELL WALL MORPHOGENESIS). ANTIFUNGAL
CC       AGENT.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- INDUCTION: SPECIFICALLY INDUCED BY CHITIN AND IS CATABOLITE
CC       REPRESSED.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; S78423; G999376; -.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ZYMOGEN;
KW   CHITIN-BINDING.
FT   SIGNAL        1     22       POTENTIAL.
FT   PROPEP       23     34
FT   CHAIN        35    423       42 KD ENDOCHITINASE.
FT   ACT_SITE    171    171       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    218    218       POTENTIAL.
SQ   SEQUENCE   423 AA;  46056 MW;  899DA50A CRC32;

  Query Match           20.6%;  Score 561;  DB 2;  Length 423;
  Best Local Similarity  35.7%;  Pred. No. 6.68e-96;
  Matches   101;  Conservative   71;  Mismatches  94;  Indels  17;  Gaps  13;

Db    111 qlfk-vkkanrglkvllsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidi 168
          | |: :|| |  || ||:|||| :| :|    :|  ||  | ::|| |::  :|||:|:
Qy     57 QEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 116

Db    169 dweyp-adatqasnmilllkevrsqrdayaaqy-apg-yhflltiaapagkdny-sklrl 224
          ||||| :::: | ::  :   |:  :|:  :  ::|  ::||: | |||     :   :
Qy    117 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 176

Db    225 adlgqvldyinlmaydyagsfspltghdanlfnnpsnpnatp-fntdsavkdyinggvpa 283
            ::| ||::|||||: ||   :|||:: |::   :: |:: :| |:|| : :  | ||
Qy    177 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 236

Db    284 nkivlgmpiygrsf--qntag--igqtyngvgs-gsw--eagiwdykalpk-agatvqyd 335
          :|::|||| |||||   :::   :|    | |: |:   |:|:  |  :    ||| |
Qy    237 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRI 296

Db    336 svakgyysynsatkelisfdtpdmintkvaylkslglggsmfw 378
            |  | :    :: ::||  : ::|||:|||  ||||:| |
Qy    297 QDQKVPYIFRD--NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337
```

```
RESULT    7
ID   CHI1_COCIM     STANDARD;      PRT;   427 AA.
AC   P54196;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   ENDOCHITINASE 1 PRECURSOR (EC 3.2.1.14) (COMPLEMENT-FIXATION ANTIGEN)
DE   (CF-ANTIGEN) (CF-AG).
GN   CTS1.
OS   COCCIDIOIDES IMMITIS.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C735;
RX   MEDLINE; 96144270.
RA   PISHKO E.J., KIRKLAND T.N., COLE G.T.;
RL   GENE 167:173-177(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SILVEIRA;
RA   YANG C., ZHU Y., MAGEE D.M., COX R.A.;
RL   SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; L41663; G1200190; -.
DR   EMBL; U51271; G1256769; -.
DR   EMBL; U33265; G1255728; -.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL;
KW   GLYCOPROTEIN.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?    427       ENDOCHITINASE 1.
FT   CARBOHYD    387    387       POTENTIAL.
FT   CONFLICT     15     47       RWLSLRCLFCELGRFMFTLSTVTAVTVVTDDIQ -> VQAS
FT                                SMSSMPNYYPVPEAPEGGFRSVVYFVNW (IN REF. 2).
FT   CONFLICT    199    199       K -> N (IN REF. 2).
SQ   SEQUENCE   427 AA;  47629 MW;  6CB9AF73 CRC32;

  Query Match           20.1%;  Score 546;  DB 2;  Length 427;
  Best Local Similarity  34.1%;  Pred. No. 1.75e-92;
  Matches    95;  Conservative   71;  Mismatches  94;  Indels  19;  Gaps  17;

Db    115 lkknnrnlktllsiggwty-spnfktpasteegrkkfadtslklmkdlgfdgididweyp 173
          ||| | :|||||:|||| : : :|    :| : |  | :::::::  :|||:|:|||||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    174 ede-kqandfvlllkacrealdaysak-hpngk-kflltiaspagpqny-nk-lklaemd 228
          :   | |   :    ::  :|:    : :|| ::||: | ||| | | :    :  :
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 180

Db    229 kyldfwnlmaydfsgswdkvsghmsnvfpsttkp-estpfssdkavkdyikagvpankiv 287
            ||| ||||||| |||:||:|| | ::      :  ::::: | || : :  | ||:|::
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    288 lgmplygraf--ast-dg-igtsfngvgg-gsw--engvwdykdmpq-qgaqvteledia 339
          |||| |||:|  ||: |  :|::  | |  |:   | |:  | ::    ||    ::|
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 300

Db    340 asysydknkrylisydtvkiagkkaeyitkngmgggm-w 377
            :| : ::::  :::| |      |  |:  :|:||:| |
Qy    301 VPYIF-RDNQW-VGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT    8
ID   CHIA_SERMA     STANDARD;      PRT;   563 AA.
AC   P07254;
DT   01-APR-1988 (REL. 07, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE A PRECURSOR (EC 3.2.1.14).
GN   CHIA.
OS   SERRATIA MARCESCENS.
```

```
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   ENTEROBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KOO J.C., LIM C.O., CHOI Y.J., KIM C.Y., BAHK J.D., LEE S.Y.,
RA   CHO M.J.;
RL   SUBMITTED (JAN-1993) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 990 / QMB1466;
RA   JONES J.D.G., GRADY K.L., SUSLOW T.V., BEDBROOK J.R.;
RL   EMBO J. 5:467-473(1986).
RN   [3]
RP   REVISIONS, AND X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE; 95219379.
RA   PERRAKIS A., TEWS I., DAUTER Z., OPPENHEIM A.B., CHET I., WILSON K.S.,
RA   VORGIAS C.E.;
RL   STRUCTURE 2:1169-1180(1994).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; L01455; G152818; -.
DR   EMBL; X03657; G46831; -.
DR   PIR; A25090; A25090.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL.
FT   SIGNAL       1     23
FT   CHAIN       24    563       CHITINASE A.
FT   DOMAIN     150    563       CATALYTIC.
FT   ACT_SITE   315    315       PROTON DONOR (PROBABLE).
FT   ACT_SITE   391    391       PROBABLE.
FT   CONFLICT    76     77       TA -> GP (IN REF. 2).
FT   CONFLICT   139    139       A -> P (IN REF. 2).
FT   CONFLICT   226    226       V -> I (IN REF. 2).
FT   CONFLICT   395    395       P -> A (IN REF. 1 AND 2).
FT   CONFLICT   410    430       PAWKPDTAYTTVNGVNALLAQ -> RPGSRHRLHHGERRQC
FT                               AAGQ (IN REF.1 AND 2).
FT   CONFLICT   437    437       V -> I (IN REF. 2).
FT   CONFLICT   464    467       ATGP -> HRA (IN REF. 2).
FT   CONFLICT   473    473       K -> E (IN REF. 2).
FT   CONFLICT   484    484       G -> S (IN REF. 2).
SQ   SEQUENCE   563 AA;  60979 MW;  7764DB57 CRC32;

  Query Match           16.1%;  Score 437;  DB 2;  Length 563;
  Best Local Similarity  28.8%;  Pred. No. 6.12e-68;
  Matches    93;  Conservative   84;  Mismatches 123;  Indels  23;  Gaps  19;

Db    231 pfaalqkaqkgvtawddpykgnfgqlmalkqahpdlkilpsiggwtlsdpfff-mgdkvk 289
          :||:: : | : | |:|     : :: :||  :| || | :|||| ::   |  |    :
Qy     36 AFAGMTNHQLSTTEWNDE-T-LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATAN 93

Db    290 -rdrfvgsvkeflqtwkffdgvdidwefpggkganpnlgspqdgetyvllmkelramldq 348
           |: || |   ||:   | ||:|:|||:||: |: | :    | | :  |: :|   ::|
Qy     94 NRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV----DKERFTTLVQDLANAFQQ 147

Db    349 lsaetgr-kyeltsaisagkdkidkvaynvaq--nsmdhiflmsydfygpfdlknlghqt 405
           :  :|: :  |::|::||   :|  :|:|     ::| : ||:||| |: : |  || :
Qy    148 EAQTSGKERLLLSAAVPAGQTYVD-AGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNS 205

Db    406 alnapawkpdtayt-tvng-vnallaqgvkpgkvvvgtamygrgwtgvngyqnnipftgt 463
          :|      : :| :  |:: |   |  |  ::|:::| : |||: |  :: :  :   :|
Qy    206 PLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPAT 265

Db    464 atg-pvkgtwkngivdyrqiagqfmsgewqytydataeapyvfkpstgdlitfddarsvq 522
          ::| |   |   |:: | :: :  :::  |   :     ||:|:    : : |||  |
Qy    266 GSGTPGPFTKEGGMLAYYEVCS-WKGATKQ-RIQD-QKVPYIFR-DN-QWVGFDDVESFK 320

Db    523 akgkyvldkqlgglfsweidadn 545
          :|  |: :| |||   | :| |:
Qy    321 TKVSYLKQKGLGGAMVWALDLDD 343


RESULT   9
ID   CHIT_NPVAC     STANDARD;      PRT;   551 AA.
AC   P41684;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   PROBABLE ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
OS   AUTOGRAPHA CALIFORNICA NUCLEAR POLYHEDROSIS VIRUS (ACMNPV).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; BACULOVIRIDAE; EUBACULOVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C6;
RX   MEDLINE; 94303173.
RA   AYRES M.D., HOWARD S.C., KUZIO J., LOPEZ-FERBER M., POSSEE R.D.;
RL   VIROLOGY 202:586-605(1994).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; L22858; G559195; -.
DR   PROSITE; PS00014; ER_TARGET.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18    551       PROBABLE ENDOCHITINASE.
FT   ACT_SITE   305    305       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD   173    173       POTENTIAL.
FT   CARBOHYD   444    444       POTENTIAL.
FT   SITE       548    551       PREVENT SECRETION FROM ER (POTENTIAL).
SQ   SEQUENCE   551 AA;  61368 MW;  BDB124DB CRC32;

  Query Match           14.5%;  Score 395;  DB 2;  Length 551;
  Best Local Similarity  29.2%;  Pred. No. 1.16e-58;
  Matches    95;  Conservative   80;  Mismatches 122;  Indels  28;  Gaps  27;

Db    221 pwaavqkpqkgvsawn-ep-ykgnfgqlmaaklanphlkilpsiggwtlsdp-fyfmhdv 277
          : |:: : | : : || |  |  :|  |:  |: || || | :|||| ::   |  |
Qy     36 AFAGMTNHQLSTTEWNDETLYQ-EFNGLK--KM-NPKLKTLLAIGGWNFGTQKFTDMVAT 91

Db    278 -ekrnvfvdsvkeflqvwkffdgvdidwefpggkganpslgdadgdaktyillleelram 336
           ::|  ||:|   ||:   | ||:|:|||:||: |: |::|   : : :  |:::|
Qy     92 ANNRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV-D-K-E-R-FTTLVQDLANA 144

Db    337 lddleaqt-grv-yeltsaisagydkiavvnya-e-aqkslgkiflmsydfkgawsntdl 392
          ::: |||| |:    |::|::|| :      | :   :|  : ||:||| |:| :
Qy    145 FQQ-EAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVT- 201

Db    393 gyqttvya-ps-wnseelytthyavdallkqgvdpnkiivgvamygrgwtgvtnytndny 450
          |  : :|       :        ||:  |  |  ::|:|:|:: |||: |  ::
Qy    202 GHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVG 261

Db    451 fsgtgng-pgsgtwedgvvdyrqiqkdlnnyvytfdsaaqasyvfdkskgdlisfdsvds 509
          ::||:| ||: | | |:: | ::          ::     :|:| : : : ::|| |:|
Qy    262 APATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQ-KVPYIF-RDN-QWVGFDDVES 318

Db    510 vlgkvkyvdrnklgglfaweidadn 534
            || |: :: |||   | :| |:
Qy    319 FKTKVSYLKQKGLGGAMVWALDLDD 343


RESULT   10
ID   CHIA_ALTSO     STANDARD;      PRT;   820 AA.
AC   P32823;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE A PRECURSOR (EC 3.2.1.14) (CHI-A).
GN   CHIA.
OS   ALTEROMONAS SP. (STRAIN O-7).
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   VIBRONACEAE.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93106952.
RA   TSUJIBO H., ORIKOSHI H., TANNO H., FUJIMOTO K., MIYAMOTO K.,
RA   IMADA C., OKAMI Y., INAMORI Y.;
RL   J. BACTERIOL. 175:176-181(1993).
RN   [2]
RP   SEQUENCE OF 22-48, AND CHARACTERIZATION.
RX   MEDLINE; 93099552.
RA   TSUJIBO H., YOSHIDA Y., MIYAMOTO K., IMADA C., OKAMI Y., INAMORI Y.;
RL   CAN. J. MICROBIOL. 38:891-897(1992).
RN   [3]
RP   MUTAGENESIS.
RX   MEDLINE; 93379363.
RA   TSUJIBO H., ORIKOSHI H., IMADA C., OKAMI Y., MIYAMOTO K.,
RA   INAMORI Y.;
RL   BIOSCI. BIOTECHNOL. BIOCHEM. 57:1396-1397(1993).
CC   -!- THE OPTIMUM PH AND TEMPERATURE OF CHI-A ARE 8.0 AND 50 DEGREES
CC       CELSIUS, RESPECTIVELY.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- ENZYME REGULATION: STIMULATED BY MG2+; INHIBITED BY
CC       N-BROMOSUCCINIMIDE AND 2-HYDROXY-5-NITROBENZYL BROMIDE.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; D13762; G216207; -.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL.
FT   SIGNAL        1     21
FT   CHAIN        22    820       CHITINASE A.
FT   ACT_SITE    313    313       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   820 AA;  87346 MW;  838F6EE8 CRC32;

  Query Match           12.7%;  Score 345;  DB 2;  Length 820;
  Best Local Similarity  35.0%;  Pred. No. 8.67e-48;
  Matches    57;  Conservative   44;  Mismatches  47;  Indels  15;  Gaps  13;

Db    250 vysqlmalkqrypdlkilpsvggwtlsdp-fhgftnkan-rdtfvasvkqflktwkfydg 307
          :| :: :||   | || | ::||| ::   |      || |:||| |  :||:   | ||
Qy     55 LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSF-DG 113

Db    308 vdidwefpggdgpnpdlgdpindgpayvalmqelramldeleaet-grq-yeltsaigag 365
          :|:|||:||::|: | : |  ::   : :|:|:|   ::: ||:| |::   |::|: ||
Qy    114 LDLDWEYPGSQGS-PAV-D--KE-R-FTTLVQDLANAFQQ-EAQTSGKERLLLSAAVPAG 166

Db    366 ydkiedvdyqaaq--qymdyifamtydfygawnnetghqtgiy 406
           :   |  |:     | :|::  |:||| |:|:: ||| : :|
Qy    167 -QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLY 208


RESULT   11
ID   CHIB_SERMA     STANDARD;      PRT;   499 AA.
AC   P11797;
DT   01-OCT-1989 (REL. 12, CREATED)
DT   01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE B PRECURSOR (EC 3.2.1.14).
GN   CHIB.
OS   SERRATIA MARCESCENS.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   ENTEROBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 990 / QMB1466;
RX   MEDLINE; 89345110.
RA   HARPSTER M.H., DUNSMUIR P.;
RL   NUCLEIC ACIDS RES. 17:5395-5395(1989).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; X15208; G47228; -.
DR   PIR; S04856; S04856.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL.
FT   SIGNAL        1     41
FT   CHAIN        42    499       CHITINASE B.
FT   ACT_SITE    144    144       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   499 AA;  55464 MW;  9A585FFC CRC32;

  Query Match           12.4%;  Score 337;  DB 2;  Length 499;
  Best Local Similarity  34.0%;  Pred. No. 4.53e-46;
  Matches    54;  Conservative   47;  Mismatches  44;  Indels  14;  Gaps  10;

Db     78 ltalkahnpslrimfsiggwyysndlgvshanyvnavktpaartkfaqscvrimkdygfd 137
          : :||  || |: :::|||||   | :| :: :: : | |:  |  |  | :|::: |:||
Qy     59 FNGLKKMNPKLKTLLAIGGW---N-FG-TQ-KFTDMVATANNRQTFVNSAIRFLRKYSFD 112

Db    138 gvdidweyp--qaa-evd-g-fiaalqeirtllnqqtiadgrqalpyqltiagaggaffl 192
          |:|:|||||  |::  ||   | : :|::   : |:: : |:: |   |: | ::|  ::
Qy    113 GLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERL--LLSAAVPAGQTYV 170

Db    193 sryysklaqivapldyinlmtydlagpwekitnhqaalf 231
              :  |   ||::|||:||: |:|||:| | ::|:
Qy    171 DAGYE-VDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLY 208


RESULT   12
ID   CHIT_STRLI     STANDARD;      PRT;   619 AA.
AC   P36909;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE C PRECURSOR (EC 3.2.1.14).
GN   CHIC.
OS   STREPTOMYCES LIVIDANS.
OC   PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93294525.
RA   FUJII T., MIYASHITA K.;
RL   J. GEN. MICROBIOL. 139:677-686(1993).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- INDUCTION: BY CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS A BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC       (CBD).
DR   EMBL; D12647; G391910; -.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS00018; EF_HAND.
DR   PROSITE; PS00561; CBD_BACTERIAL.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    619       CHITINASE C.
FT   DOMAIN       35    140       CELLULOSE-BINDING.
FT   DOMAIN      148    230       FIBRONECTIN TYPE-III.
FT   DOMAIN      240    619       CATALYTIC.
FT   ACT_SITE    382    382       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   619 AA;  65200 MW;  23E3E1EC CRC32;

  Query Match           11.1%;  Score 301;  DB 2;  Length 619;
  Best Local Similarity  28.2%;  Pred. No. 2.01e-38;
  Matches    86;  Conservative   65;  Mismatches 129;  Indels  25;  Gaps  18;

Db    311 w-dqplrgnfnqlrklkakypnikilysfggwtwsgg-fpdavknpaafakschdlvedp 368
          | |: |  :|| |:|::   |::| | ::|||  :   | | |  :       :  :
Qy     50 WNDETLYQEFNGLKKMN---PKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFL 106

Db    369 rwadvfdgidldweypnacglscdetsapnafssmmkamraefgqdylita-a---vtad 424
          |    |||:|||||| : | |   :   : |:::: :   | |:    ::     ::|
Qy    107 RKYS-FDGLDLDWEYPGSQG-S--PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAA 162
```

```
Db    425 gsdggk-idaa-dygeaskyidwynvmtydffgawak-ngptaphspltaydgipqqgfn 481
          : |   :||: :    :  :|  |:|:||| |:| |  |  :|       | :  ::|
Qy    163 VPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA-SLN 221

Db    482 tadamakfkskgvpadklligigfygrgwt-gvtqsap-ggtatgpat-gty--eagied 536
            |:     || || ||::|:  |||: | : :  :  |: |||::| | :  |:|:
Qy    222 VDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLA 281

Db    537 y-kvlknscpatgtiag--tayahcgsnwwsydtpatikskmdwaeqqglggaffwefsg 593
          | |     ::   |     :|    : | ::|   ::|:|:    | |||||  | :
Qy    282 YYEVCSWKGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDL 341

Db    594 dtang 598
          | |
Qy    342 DDFAG 346


RESULT   13
ID   CHIT_STRPL     STANDARD;      PRT;   610 AA.
AC   P11220;
DT   01-JUL-1989 (REL. 11, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE 63 PRECURSOR (EC 3.2.1.14).
GN   CHTA.
OS   STREPTOMYCES PLICATUS.
OC   PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92192480.
RA   ROBBINS P.W., OVERBYE K., ALBRIGHT C., BENFIELD B., PERO J.;
RL   GENE 111:69-76(1992).
RN   [2]
RP   SEQUENCE OF 1-45 FROM N.A., AND SEQUENCE OF 31-45.
RX   MEDLINE; 88087127.
RA   ROBBINS P.W., ALBRIGHT C., BENFIELD B.;
RL   J. BIOL. CHEM. 263:443-447(1988).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- INDUCTION: BY CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS A BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC       (CBD).
DR   EMBL; M82804; G153216; -.
DR   EMBL; M18397; G153209; -.
DR   PIR; A29912; A29912.
DR   PIR; JH0573; JH0573.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS00018; EF_HAND.
DR   PROSITE; PS00561; CBD_BACTERIAL.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL.
FT   SIGNAL        1     30
FT   CHAIN        31    610       CHITINASE 63.
FT   DOMAIN       35    140       CELLULOSE-BINDING.
FT   DOMAIN      148    229       FIBRONECTIN TYPE-III.
FT   DOMAIN      236    610       CATALYTIC.
FT   ACT_SITE    383    383       PROTON DONOR (BY SIMILARITY).
FT   CONFLICT      3      3       F -> I (IN REF. 2).
SQ   SEQUENCE   610 AA;  63974 MW;  2F5E8E35 CRC32;

  Query Match           10.5%;  Score 287;  DB 2;  Length 610;
  Best Local Similarity  26.4%;  Pred. No. 1.72e-35;
  Matches    78;  Conservative   63;  Mismatches 134;  Indels  21;  Gaps  16;

Db    320 fnqlrnlkaeyphikilysfggwtwsgg-fpdavknpaafakschdlvedprwadvfdgi 378
          : ::  ||   | :| | ::|||  :   | | |  :       :  :   |    |||:
Qy     56 YQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYS-FDGL 114

Db    379 dldweypnacglscdetsapnafssmmkamraefgqdylita-a---vtadgsdggk-id 433
          |||||| : | |   :   : |::::  :   | |:   ::      ::|  : |   :|
Qy    115 DLDWEYPGSQG-S--PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVD 171

Db    434 aa-dygeaskyidwynvmtydffgawak-ngptaphsplnaydgipqqgfttadamakfk 491
          |: :    :  :|  |:|:||| |:| |  |  :|       | :  ::    |:
Qy    172 AGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA-SLNVDAAVQQWL 230

Db    492 skgvpadklligigfygrgwt-gvtqsap-ggtatgpaa-gty--eagiedy-kvlknsc 545
          || || ||::|:  |||: | : :  :  |: |||::: | :  |:|:  |  |
Qy    231 QKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKG 290

Db    546 patgtvag--tayahcgtnwwsydtpatikskmdwaeqqglggaffwefsgdttng 599
          ::   :     :|      | ::|   ::|:|:    | |||||  | :  |   |
Qy    291 ATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAG 346


RESULT   14
ID   KTXA_KLULA     STANDARD;      PRT;  1146 AA.
AC   P09805;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   KILLER TOXIN ALPHA AND BETA SUBUNITS PRECURSOR (RF2 PROTEIN)
DE   (ENDOCHITINASE (EC 3.2.1.14)).
OS   KLUYVEROMYCES LACTIS (YEAST).
OG   PLASMID PGKL-1.
OC   EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS 2359/152;
RA   SOR F., FUKUHARA H.;
RL   CURR. GENET. 9:147-155(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84297209.
RA   STARK M.J.R., MILEHAM A.J., ROMANOS M.A., BOYD A.;
RL   NUCLEIC ACIDS RES. 12:6011-6030(1984).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85037931.
RA   TOKUNAGA M., KAWAMURA A., HISHINUMA F.;
RL   NUCLEIC ACIDS RES. 12:7581-7597(1984).
RN   [4]
RP   IDENTIFICATION OF PROTEIN, AND SEQUENCE OF 30-44 AND 895-916.
RX   MEDLINE; 87004569.
RA   STARK M.J.R., BOYD A.;
RL   EMBO J. 5:1995-2002(1986).
RN   [5]
RP   SIMILARITY TO CHITINASE OF ALPHA-SUBUNIT.
RX   MEDLINE; 90259069.
RA   BRADSHAW H.D. JR.;
RL   NATURE 345:299-299(1990).
RN   [6]
RP   CHITINASE ACTIVITY OF ALPHA-SUBUNIT.
RX   MEDLINE; 91301161.
RA   BUTLER A.R., O'DONNEL R.W., MARTIN V.J., GOODAY G.W., STARK M.J.R.;
RL   EUR. J. BIOCHEM. 199:483-488(1991).
CC   -!- FUNCTION: THE ALPHA SUBUNIT IS A POTENT EXOCHITINASE. ALONG WITH
CC       THE BETA SUBUNIT IT PLAYS A ROLE IN THE INITIAL INTERACTION OF THE
CC       TOXIN WITH SENSITIVE CELLS AND ALLOW THE GAMMA SUBUNIT (THE ACTIVE
CC       TOXIN) TO GAIN ENTRY INTO THE CELL.
CC   -!- PTM: RF2 IS POTENTIALLY SPLIT BY MEMBRANE-BOUND BASIC AMINO ACID-
CC       SPECIFIC PEPTIDASE TO YIELD THE ALPHA AND BETA SUBUNITS.
CC   -!- SUBUNIT: THE KILLER TOXIN IS COMPOSED OF THREE SUBUNITS: ALPHA,
CC       BETA AND GAMMA.
CC   -!- SIMILARITY: LOCAL, TO OTHER CHITIN-BINDING PROTEINS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
DR   EMBL; X07127; G2829; -.
DR   EMBL; X00762; G2844; -.
DR   EMBL; X01095; G2849; -.
DR   PIR; S07915; S07915.
DR   HSSP; P02877; 1HEV.
DR   PROSITE; PS00026; CHITIN_BINDING.
```

```
DR   PROSITE; PS01095; CHITINASE_18.
KW   PLASMID; TOXIN; GLYCOPROTEIN; SIGNAL; HYDROLASE; GLYCOSIDASE;
KW   CHITIN DEGRADATION.
FT   SIGNAL        1     17       POTENTIAL.
FT   PROPEP       18     29       POTENTIAL.
FT   CHAIN        30    892       ALPHA SUBUNIT (PROBABLE).
FT   CHAIN       895   1146       BETA SUBUNIT (PROBABLE).
FT   DOMAIN      310    363       CHITIN-BINDING (BY SIMILARITY).
FT   DOMAIN      364    635       CATALYTIC.
FT   ACT_SITE    495    495       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    771    771       POTENTIAL.
FT   CARBOHYD    858    858       POTENTIAL.
FT   CARBOHYD    868    868       POTENTIAL.
FT   CARBOHYD    876    876       POTENTIAL.
FT   CARBOHYD    995    995       POTENTIAL.
FT   CARBOHYD   1082   1082       POTENTIAL.
FT   CARBOHYD   1117   1117       POTENTIAL.
SQ   SEQUENCE   1146 AA;  128936 MW;  CDD1620F CRC32;

  Query Match             9.7%;  Score 263;  DB 5;  Length 1146;
  Best Local Similarity  28.4%;  Pred. No. 1.56e-30;
  Matches    60;  Conservative   58;  Mismatches  74;  Indels  19;  Gaps  14;

Db    425 ddsafsksaflkvtssk-k-ipsfggwdfstspstytifrnavktdqnrntfannlinfm 482
          |:: :     ||  ::| | : ::|||:|:|    :|   : | |  || || |: | |:
Qy     52 DETLYQEFNGLKKMNPKLKTLLAIGGWNFGT-Q-KFT---DMVATANNRQTFVNSAIRFL 106

Db    483 nkynldgidldweypgapdipdipaddsssgsnyltflkllkgkmps-gktl-siaipss 540
           ||::||:|||||||||:   | :  :  ::    |:      :  ::  : | | |:|::
Qy    107 RKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG 166

Db    541 ywylkn-fpisdiqntvdymvymtydihgiwe--ygkans-yi---nchtprk-eiedai 592
            |:   : :  |   :|::  |:||:|| ||   |  :: |    :  ::   ::: |:
Qy    167 QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAV 226

Db    593 km-ldkagvkfnkvfggvanygrsykmvntn 622
             |:| |   :|:: |:: |||| : :::
Qy    227 QQWLQK-GTPASKLILGMPTYGRSFTLASSS 256


RESULT   15
ID   GP39_BOVIN     STANDARD;      PRT;    36 AA.
AC   P30922;
DT   01-JUL-1993 (REL. 26, CREATED)
DT   01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   CARTILAGE GLYCOPROTEIN-39 (GP-39) (39 KD WHEY PROTEIN) (FRAGMENT).
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; ARTIODACTYLA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 88106603.
RA   REJMAN J.J., HURLEY W.L.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 150:329-334(1988).
CC   -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE CAPACITY OF CELLS TO
CC       RESPOND TO AND COPE WITH CHANGES IN THEIR ENVIRONMENT.
CC   -!- PTM: GLYCOSYLATED.
CC   -!- TISSUE SPECIFICITY: MAMMARY SECRETIONS COLLECTED DURING THE
CC       NONLACTATING PERIOD.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
DR   PIR; A27682; A27682.
KW   GLYCOPROTEIN.
FT   NON_TER      36     36
SQ   SEQUENCE   36 AA;  4264 MW;  F3935740 CRC32;

  Query Match             6.3%;  Score 172;  DB 4;  Length 36;
  Best Local Similarity  48.6%;  Pred. No. 8.27e-13;
  Matches    17;  Conservative   11;  Mismatches   7;  Indels   0;  Gaps   0;

Db      2 klixyytswsqyregdgsxfpdaidpflxthviys 36
          ||: |:|:|:|||:|::  :|  :|| | ||:||:
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYA 36


Search completed: Fri Jun 27 17:00:43 1997
Job time : 63 secs.
```

****

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 27 17:08:30 1997; MasPar time 9.60 Seconds
423.400 Million cell updates/sec

Tabular output not generated.

Title: >US-08-663-618A-15
Description: (1-373) from US08663618A.pep
Perfect Score: 2717
Sequence: 1 AKLVCYFTNWAQYRQGEARF.....YPLIQTLRQELSLPYLSSGT 373

Scoring table: PAM 150
Gap 11

Searched: 92623 seqs, 10896596 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq26
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19

Statistics: Mean 34.957; Variance 153.180; scale 0.228

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1328	48.9	537	14	R73991	Bovine oviduct specif	9.77e-114
2	1303	48.0	668	14	R73993	Hamster oviduct speci	2.57e-111
3	1303	48.0	718	14	R73992	Murine oviduct specif	2.57e-111
4	881	32.4	321	1	P81342	Polypeptide involved	1.13e-70
5	597	22.0	371	13	R70025	Tobacco chitinase enc	1.07e-43
6	597	22.0	377	13	R70029	Tobacco cluster-A pro	1.07e-43
7	581	21.4	424	12	R63441	Trichoderma harzianum	3.42e-42
8	576	21.2	389	10	R56861	Aphanocladium album m	1.01e-41
9	576	21.2	423	6	R33072	Aphanocladium album p	1.01e-41
10	569	20.9	389	6	R33068	Aphanocladium album m	4.58e-41
11	463	17.0	866	19	W02159	Soluble chitinase.	3.52e-31
12	389	14.3	561	12	R64823	Chitinase A.	2.29e-24
13	372	13.7	561	9	R48981	Bacterial chitinase-A	8.14e-23
14	271	10.0	238	6	R32547	HaNPV IE-1 gene prote	9.44e-14
15	219	8.1	1046	19	W02156	Periplasmic chitodext	3.13e-09
16	139	5.1	25	13	R70751	YKL-40 N-terminal pep	1.22e-02
17	139	5.1	25	12	R70745	YKL-40 N-terminal seq	1.22e-02
18	136	5.0	29	11	R60073	Bovine Oviduct Glycop	2.09e-02
19	101	3.7	1070	4	R21521	Alpha galactosidase.	8.48e+00
20	99	3.6	398	18	R95856	S. pyogenes speB gene	1.17e+01
21	99	3.6	428	16	R86905	Murine BMP-9.	1.17e+01
22	99	3.6	428	6	R31372	Murine bone morphogen	1.17e+01
23	99	3.6	553	14	R80445	Murine soluble epoxid	1.17e+01
24	99	3.6	958	1	P81181	Sequence of glucoamyl	1.17e+01
25	99	3.6	958	2	R07575	Glucoamylase encoded	1.17e+01
26	95	3.5	438	9	R47259	Pre-pro-VGR1.	2.23e+01
27	92	3.4	1356	5	R26999	Novel type III RTK en	3.58e+01
28	91	3.3	19	13	R70752	YKL-40 internal pepti	4.18e+01
29	91	3.3	19	12	R70746	YKL-40 internal pepti	4.18e+01
30	90	3.3	256	17	R97235	MuTu putative oncogen	4.89e+01
31	90	3.3	459	17	R88058	Protein encdoed by Mu	4.89e+01
32	90	3.3	575	3	P50018	Sequence encoded by t	4.89e+01
33	90	3.3	709	2	P70423	FixD nitrogen fixatio	4.89e+01
34	90	3.3	709	3	P50017	Sequence encoded by t	4.89e+01
35	87	3.2	369	18	R98125	Human lymphocyte cell	7.77e+01
36	88	3.2	513	9	R47264	Pre-pro BMP6.	6.66e+01
37	88	3.2	513	2	R07311	Human Bone Morphogene	6.66e+01
38	88	3.2	513	6	R32904	Human BMP-6 peptide s	6.66e+01
39	88	3.2	513	7	R36735	Human BMP-6.	6.66e+01
40	87	3.2	542	16	R90772	Bacillus stearothermo	7.77e+01
41	88	3.2	896	17	R92750	Human EGF receptor su	6.66e+01
42	86	3.2	2332	8	R43257	Human Factor VIII.	9.05e+01
43	87	3.2	2351	3	P50319	Human antihaemophilic	7.77e+01
44	86	3.2	2351	3	P50059	Human factor VIII.	9.05e+01
45	86	3.2	2351	13	R78223	Human Factor-VIII:c.	9.05e+01

ALIGNMENTS

RESULT 1
ID R73991 standard; Protein; 537 AA.
AC R73991;
DT 10-JAN-1996 (first entry)
DE Bovine oviduct specific glycoprotein.
KW Bovine oviduct specific glycoprotein; recombinant production;
KW BOGP.
OS Bos taurus.
FH Key Location/Qualifiers
FT Peptide 1..18
FT /label= sig_peptide
FT Peptide 19..537
FT /label= mat_peptide
PN J07107979-A.
PD 25-APR-1995.
PF 15-AUG-1994; 214227.
PR 19-AUG-1993; JP-227881.
PA (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR WPI; 95-190179/25.
DR N-PSDB; Q90442.
PT New DNA encoding an oviduct-specific glycoprotein - useful for
PT recombinant protein production in high quantities.
PS Claim 4; Pages 7-9; 22pp; Japanese.
CC Q90442 encodes R73991 bovine oviduct specific glycoprotein (BOGP).
CC The cDNA can be used for the commercial recombinant prodn. of
CC BOGP in high quantities.
SQ Sequence 537 AA;

Query Match 48.9%; Score 1328; DB 14; Length 537;
Best Local Similarity 51.2%; Pred. No. 9.77e-114;
Matches 187; Conservative 74; Mismatches 97; Indels 7; Gaps 6;

```
Db     20 klvcyftnwafsrpgpasilprdldpflcthlvfafasmsnnqivpkdpqdekilypefn 79
          ||||||||||  | | | :||:|||| ||||||::|||:|:|:|:   :  ||  || |||
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDET-LYQEFN 60

Db     80 klkernrglktllsiggwnfgtvrfttmlstfsnrerfvssviallrthgfdgldlffly 139
             ||  |  |||||:|||||||| :|| |::| :||: ||:| | :||  :||||||   |
Qy     61 GLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEY 120

Db    140 pglrgspardrwtfvflleellqafkneaqltmrprlllsaavsgdphvvqkayearllg 199
            || :|||| |:  |  |:::|  ||  ||| : : ||||||||::    |: :||   ::
Qy    121 PGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIA 180
```

```
Db    200 rlldfisvlsydlhgswekvtghnsplfs-l--pgdpks-sa-yamnywrqlgvppekll 254
          : |||:::::||:|||||||||||||||:     :| : | :   |:  | | | |: ||:
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    255 mglptygrtfhllkasqnelraqavgpaspgkytkqagflayyeiccfvrrakkrwindq 314
          :|:|||||:| |  :|:  : | | |:::|| :||::| |||||:|   : | |::| ||
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQ 299

Db    315 yvpyafkgkewvgyddaisfgykaffikrehfggamvwtldlddfrgyfcgtgpfplvht 374
            ||| |: ::|||:|| || | ::|:  :||||||:|||||| |: | | :||::|
Qy    300 KVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYPLIQT 359

Db    375 lnnll 379
            | |
Qy    360 LRQEL 364


RESULT    2
ID   R73993 standard; Protein; 668 AA.
AC   R73993;
DT   10-JAN-1996  (first entry)
DE   Hamster oviduct specific glycoprotein.
KW   Hamster oviduct specific glycoprotein; recombinant production;
KW   HOGP.
OS   Cricetulus griseus.
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT   /label= sig_peptide
FT   Peptide         19..668
FT   /label= mat_peptide
PN   J07107979-A.
PD   25-APR-1995.
PF   15-AUG-1994; 214227.
PR   19-AUG-1993; JP-227881.
PA   (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR   WPI; 95-190179/25.
DR   N-PSDB; Q90444.
PT   New DNA encoding an oviduct-specific glycoprotein - useful for
PT   recombinant protein production in high quantities.
PS   Claim 4; Pages 16-18; 22pp; Japanese.
CC   Q90444 encodes R73993 hamster oviduct specific glycoprotein (HOGP).
CC   The cDNA can be used for the commercial recombinant prodn. of
CC   HOGP in high quantities.
SQ   Sequence   668 AA;

  Query Match           48.0%;  Score 1303;  DB 14;  Length 668;
  Best Local Similarity  48.4%;  Pred. No. 2.57e-111;
  Matches   177;  Conservative   81;  Mismatches  99;  Indels   9;  Gaps   7;

Db     20 klvcyftnwahsrpvpasilprdldpflcthlifafasmsnnqivannlqdekilypefn 79
          ||||||||||: |   | :||:|||| ||||||:|||:|:|:|: : :  ||  || |||
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDET-LYQEFN 60

Db     80 klkernralktllsvggwnfgtsrfttmlstlasrekfigsvvsflrthgfdgldlffly 139
            ||  |  ||||::|||||| :|| |::|  :|: |: | : |||  :||||||   |
Qy     61 GLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEY 120

Db    140 pglrgspindrwnflflieelqfafekealltqrprlllsaav-sgipyiiqtsydvhll 198
          || :|||  |:  |  |:::|  ||: ||  : : |||||||| :|  |: :::|:|  :
Qy    121 PGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYV-DAGYEVDKI 179

Db    199 grrldfinvlsydlhgswekstghnsplfslpedpkssa-famn--y--wrnlgapadkl 253
          :: |||:|:::||:|||||| |||||||:   |:: ::| : ::     |   |:|| ||
Qy    180 AQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKL 239

Db    254 lmgfpaygrtfhllresknglqaasmgpaspgkytkqagflayyevcsfiqraekhwidh 313
          ::| |:|||:| |   |   : |:: |:::|| :||::| |||||||    | |::|:
Qy    240 ILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQD 298

Db    314 qyvpyaykgkewvgyddavsfsykamfvkkehfggamvwtldmddvrgtfcgngpfplvh 373
          | ||| :: ::|||:||  ||  |  ::|   :||||||:||:||  |  |  | :||::
Qy    299 QKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYPLIQ 358

Db    374 ilnell 379
            | : |
Qy    359 TLRQEL 364


RESULT    3
ID   R73992 standard; Protein; 718 AA.
AC   R73992;
DT   10-JAN-1996  (first entry)
DE   Murine oviduct specific glycoprotein.
KW   Murine oviduct specific glycoprotein; recombinant production;
KW   MOGP.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT   /label= sig_peptide
FT   Peptide         19..718
FT   /label= mat_peptide
PN   J07107979-A.
PD   25-APR-1995.
PF   15-AUG-1994; 214227.
PR   19-AUG-1993; JP-227881.
PA   (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR   WPI; 95-190179/25.
DR   N-PSDB; Q90443.
PT   New DNA encoding an oviduct-specific glycoprotein - useful for
PT   recombinant protein production in high quantities.
PS   Claim 4; Pages 11-14; 22pp; Japanese.
CC   Q90443 encodes R73992 murine oviduct specific glycoprotein (MOGP).
CC   The cDNA can be used for the commercial recombinant prodn. of
CC   MOGP in high quantities.
SQ   Sequence   718 AA;

  Query Match           48.0%;  Score 1303;  DB 14;  Length 718;
  Best Local Similarity  49.9%;  Pred. No. 2.57e-111;
  Matches   182;  Conservative   76;  Mismatches 100;  Indels   7;  Gaps   4;

Db     20 klvcyftnwahsrpgpasimphdldpflcthlifafasmsnnqivaknlqdenvlypefn 79
          ||||||||||: | | | ::| |||| ||||||:|||:|:|:|: : :  ||  || |||
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDET-LYQEFN 60

Db     80 klkernrelktllsiggwnfgtsrftamlstlanrekfidsvisflrihgfdgldlffly 139
            ||  |  ||||:|||||||| :|| |::|  ||: |::| | |||  :||||||   |
Qy     61 GLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEY 120

Db    140 pglrgspphdrwnflflieelqfaferealltqhprlllsaavsgipsiihtsydalllg 199
          || :|||: |:  |  |:::|  ||::||  :   |||||||::  : : ::|:   ::
Qy    121 PGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIA 180

Db    200 rrldfinvlsydlhgswekftghnsplfslpedskssa--y---amnywrklgtpadkli 254
          : |||:|:::||:|||||| |||||||:   |:| ::|      |:  |   |||| |||
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    255 mgfptygrnfyllkesknglqtasmgpaspgkytkqagflayyevcsfvqrakkhwidyq 314
          :| |||||:| |   |   : ::: |:::|| :||::| ||||||||    | |::|: |
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQ 299

Db    315 yvpyafkgkewlgyddtisfsykamyvkrehfggamvwtldmddvrgtfcgngpfplvhi 374
            ||| |: ::|:|:||  ||  |  |:|:  :||||||:||:||  |  |  | :||::
Qy    300 KVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYPLIQT 359

Db    375 lnell 379
            | : |
Qy    360 LRQEL 364


RESULT    4
ID   P81342 standard; protein; 321 AA.
AC   P81342;
DT   19-OCT-1990  (first entry)
DE   Polypeptide involved in protective mechanisms
KW   Immune response; cell growth.
PN   J63032898-A.
```

```
PD   30-JAN-1988.
PF   16-JUL-1986; 167518.
PR   16-JUL-1986; JP-167518.
PA   (DAIN) Dainippon Pharm KK.
DR   WPI; 88-068419/10.
DR   P-PSDB; P81342.
PT   New polypeptide and DNA encoding it - related to protective
PT   mechanisms such as immune response etc.
PS   Disclosure; p; Japanese.
CC   This polypeptide is involved in protective mechanisms such as immune
CC   response, cell growth and activation of protective functions.
SQ   Sequence   321 AA;

  Query Match           32.4%;  Score 881;  DB 1;  Length 321;
  Best Local Similarity  50.2%;  Pred. No. 1.13e-70;
  Matches   108;  Conservative   60;  Mismatches  41;  Indels   6;  Gaps   3;

Db     22 klvcyytswsqyregdgscfpdaldrflcthiiysfanisndhidtwewndvtlygmlnt 81
          |||||:|:|:|||:|::  :|  ||  ||||:||:|| ::| :: | |||| |||  :|
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDETLYQEFNG 61

Db     82 lknrnpnlktllsvggwnfgsqrfskiasntqsrrtfiksvppflrthgfdgvdlawlyp 141
          ||: ||:|||||::|||||:|:|: : : : :|:||::|   |||  :|||:|| | ||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    142 g-rg----dkqhfttlikemkaefikeaq-pgkkqlllsaalsagkvtidssydiakisq 195
          | :|    ||::|||| ::   |  ||| :|| :|||||::||   :|::|:: ||:|
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQ 181

Db    196 hldlisimtydfhgawrgttghhsplfrgqedasp 230
          :||::::|:|||||:|   |||:|||:: ||:::
Qy    182 NLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGA 216


RESULT    5
ID   R70025 standard; Protein; 371 AA.
AC   R70025;
DT   29-SEP-1995  (first entry)
DE   Tobacco chitinase encoded by tobacco Cluster-A cDNA clone cA-3.
KW   Tobacco; chitinase; antifungal; fungicide; Cluster-A.
OS   Nicotiana tabacum.
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT   /label= signal peptide
FT   Misc_difference 22
FT   /note= "not present in genomic clone"
FT   Misc_difference 182..195
FT   /note= "Determined by protein sequencing"
FT   Misc_difference 216..241
FT   /note= "Determined by protein sequencing"
PN   EP-639642-A.
PD   22-FEB-1995.
PF   17-AUG-1993; 202425.
PR   17-AUG-1993; EP-202425.
PA   (MOGE-) MOGEN INT NV.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PI   Apotheker-de Groot M,  Bol JF,  Cornelissen BJC,  Linthorst HJM;
PI   Melchers LS,  Ponstein AS,  Sela-buurlage MB;
DR   WPI; 95-083454/12.
DR   N-PSDB; Q82973.
PT   New plant protein having endo-chitinase activity - used in
PT   antifungal compsns. and to develop transformed plants which are
PT   less susceptible to fungal infection.
PS   Claim 3; Page 19-20; 43pp; English.
CC   Screening of a lambda ZAP cDNA library of TMV-infected Samsunn NN
CC   tobacco plants with a probe derived from PROB40 (a partial Cluster-A
CC   cDNA clone) resulted in the isolated of 11 positively hybridising
CC   clones. Analysis revealed that all were identical and corresp. to
CC   Cluster-A cDNA. The nt sequence of clone cA-3 is given in Q82973/
CC   R70025. This clone lacks 7 codons of the 5' part of the ORF. The
CC   cDNA clone was completed by performing a PCR reaction on clone cA-3
CC   with primer T7 (Q82794) and primer P1 (Q82975). A partial AA
CC   sequence of the protein isolated from TMV-infected tobacco leaves
CC   was obtd. and found to be almost identical to the AA sequence
CC   deduced from the cDNA.
SQ   Sequence   371 AA;

  Query Match           22.0%;  Score 597;  DB 13;  Length 371;
  Best Local Similarity  33.0%;  Pred. No. 1.07e-43;
  Matches   107;  Conservative   86;  Mismatches 108;  Indels  23;  Gaps  18;

Db     36 nnidstlfthlfcafadlnpqsnqliispenqds-fsqftstvqrknpsvktflsiaggr 94
          :::|::| |||: ||| :   ::||  :  |::: : :| : : ::|| :||:|:| ||:
Qy     23 KDLDPSLCTHLIYAFAGMT--NHQLSTTEWNDETLYQEFNG-LKKMNPKLKTLLAI-GGW 78

Db     95 a-dttaygimarqpnsrksfidssirlarqfgfhgldldweyp--l-s-atdmtnlgill 149
          |  :  |   :|:| :|::|:||: | ::| |||||||||    | | |:  :  |:
Qy     79 NFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLV 138

Db    150 newrtainmearnsgraallltaavsysprvnglnypvesvarnlnwinlmaydfygpnw 209
          :   |:  ||: ||:  |||:|||: :       | |: :|:||: :|||||| |: |
Qy    139 QDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGS-W 197

Db    210 spsqtnshaqlfdpvnh--isgs---dg-inawiqagvptkklvlgipfygyawrlvnpn 263
          |  :: |:   :    ::|   |: :  |:| | |: ||:||:| || :  | :::
Qy    198 EKV-TGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSS 256

Db    264 ihdlrapaagksnvga-vddgsm-tynrirdyivqtrattvynativgdycysgsnwisy 321
          : |||:| :  |:   :|:| :|  :      |:   :  :   |  | :  : |:::
Qy    257 DTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQR-IQDQK-VP-YIFRDNQWVGF 313

Db    322 ddtqsvrnkvnyvkgrgllgyfaw 345
          || :| : ||:|:| :|| |   |
Qy    314 DDVESFKTKVSYLKQKGLGGAMVW 337


RESULT    6
ID   R70029 standard; Protein; 377 AA.
AC   R70029;
DT   29-SEP-1995  (first entry)
DE   Tobacco cluster-A protein encoded by genomic clone.
KW   Tobacco; chitinase; antifungal; fungicide; Cluster-A.
OS   Nicotiana tabacum.
FH   Key             Location/Qualifiers
FT   Peptide         1..25
FT   /label= signal peptide
FT   Misc_difference 28..29
FT   /note= "cDNA encodes additional Ile here"
FT   Misc_difference 188..201
FT   /label= Determined by sequencing the protein
FT   Misc_difference 222..247
FT   /note= "Determined by sequencing the protein"
PN   EP-639642-A.
PD   22-FEB-1995.
PF   17-AUG-1993; 202425.
PR   17-AUG-1993; EP-202425.
PA   (MOGE-) MOGEN INT NV.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PI   Apotheker-de Groot M,  Bol JF,  Cornelissen BJC,  Linthorst HJM;
PI   Melchers LS,  Ponstein AS,  Sela-buurlage MB;
DR   WPI; 95-083454/12.
DR   N-PSDB; Q82976.
PT   New plant protein having endo-chitinase activity - used in
PT   antifungal compsns. and to develop transformed plants which are
PT   less susceptible to fungal infection.
PS   Claim 3; Page 25-28; 43pp; English.
CC   Screening of a lambda ZAP cDNA library of TMV-infected Samsunn NN
CC   tobacco plants with a probe derived from PROB40 (a partial Cluster-A
CC   cDNA clone) resulted in the isolated of 11 positively hybridising
CC   clones. Analysis revealed that all were identical and corresp. to
CC   Cluster-A cDNA.The sequence of cDNA clone cA-3 is given in Q82973/
CC   R70025. A genomic library of N. tabacum was screened using the
CC   Cluster-A cDNA insert of clone cA-3 as a probe (see Q82977, Q82978).
CC   The complete nt. sequence of Cluster-A cDNA including the deduced
CC   primary structure of the Cluster-A protein the 5' and 3' UTR regions
CC   of the gene are shown in Q82976/R70029. Comaprison of the cDNA
```

```
CC   clone with the Cluster-A gene revealed that these sequences share
CC   a high degree of identity (94%). The Cluster-A precursor protien
CC   contains a putative signal peptide as well as 4 potential N-linked
CC   glycosylation sites (N-X-S/T). The predicted mature protein has
CC   a calculated mol. wt. of 39,033 Da.
SQ   Sequence   377 AA;

  Query Match           22.0%;  Score 597;  DB 13;  Length 377;
  Best Local Similarity  33.0%;  Pred. No. 1.07e-43;
  Matches   107;  Conservative   88;  Mismatches 106;  Indels  23;  Gaps  18;

Db     42 nnidstlfthlfcafadlnpqlnqliispenqds-frqftstvqrknpsvktflsiaggr 100
          :::|::| |||: ||| :  :  ||  :  |::: :::| : : ::|| :||:|:| ||:
Qy     23 KDLDPSLCTHLIYAFAGMTNH--QLSTTEWNDETLYQEFNG-LKKMNPKLKTLLAI-GGW 78

Db    101 ansta-ygimarqpnsrksfidssirlarqlgfhgldldweyp--l-saa-dmtnlgtll 155
            :|  :  |   :|:| :|::|:||: |  :| |||||||||    |:| |:  : ||:
Qy     79 NFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLV 138

Db    156 newrtaintearnsgraallltaavsnsprvnglnypveslarnldwinlmaydfygpnw 215
             :   |:  ||: ||:  |||:|||: :       | |: :|:||| :|||||| |: |
Qy    139 QDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGS-W 197

Db    216 spsqtnshaqlfdpvnh--vsgs---dg-inawiqagvptkklvlgipfygyawrlvnan 269
              |  :: |:   :    ::|   |: :  |:| | |: ||:||:| || :  | :::
Qy    198 EKV-TGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSS 256

Db    270 ihglrapaagksnvga-vddgsm-tynrirdyivesrattvynativgdycysgsnwisy 327
             : |||:| :  |:   :|:| :|  :       ::   : :   |  | :  : |:::
Qy    257 DTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQR-IQDQK-VP-YIFRDNQWVGF 313

Db    328 ddtqtvrnkvnyvkgrgllgyfaw 351
          || :: : ||:|:| :|| |   |
Qy    314 DDVESFKTKVSYLKQKGLGGAMVW 337


RESULT    7
ID   R63441 standard; Protein; 424 AA.
AC   R63441;
DT   06-JUL-1995  (first entry)
DE   Trichoderma harzianum P1 endochitinase.
KW   Trichoderma harzianum P1; endochitinase; inhibition;
KW   chitin containing fungi; herbivorous insects.
OS   Trichoderma harzianum.
PN   WO9424288-A.
PD   27-OCT-1994.
PF   10-FEB-1994; U01198.
PR   14-APR-1993; US-045269.
PA   (CORR ) CORNELL RES FOUND INC.
PI   Harman GE,  Hayes CK,  Lorito M,  Tronsmo A;
DR   WPI; 94-341867/42.
DR   N-PSDB; Q78144.
PT   Trichoderma harzianum P1 endo:chitinase and chitobiase genes -
PT   useful in biological control of chitin-contg. fungi and insects
PT   (Eng).
PS   Example 8; Page 45; 62pp; English.
CC   Q78144 encodes R63441 the Trichoderma harzianum P1 endochitinase,
CC   which can be used to inhibit chitin containing fungi and chitin
CC   containing herbivorous insects, especially from the genera below
CC   Fusarium, Gliocladium, Rhizoctonia, Trichoderma, Uncinula, Ustilage,
CC   Erysiphe, Botrytis, Saccharomyces, Sclerotium and Alternaria.
SQ   Sequence   424 AA;

  Query Match           21.4%;  Score 581;  DB 12;  Length 424;
  Best Local Similarity  36.0%;  Pred. No. 3.42e-42;
  Matches   102;  Conservative   69;  Mismatches 95;  Indels 17;  Gaps  13;

Db    112 qlfk-lkkangnlkvmlsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidv 169
          | |: ||| | :|| :|:|:|||  :| :|    :|  ||  | ::|| |::  :|||:|:
Qy     57 QEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 116

Db    170 dweyp-addtqatnmvlllkeirsqldayaaqy-apg-yhfllsiaapagpehy-sflhm 225
          ||||| :: : | ::  :   ::   :|:  :  ::|  ::||| | |||     :   :
```

```
Qy    117 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 176

Db    226 sdlgqvldyvnlmaydyagswssysghdanlfanpsnpnssp-yntdqaikdyikggvpa 284
            ::| ||:|||||||: |||   :||:: |:    :: :::  | | |: : :  | ||
Qy    177 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 236

Db    285 skivlgmpiygraf--estgg--iggtysgigs-gsw--engiwdy-kvlpkagatvqyd 336
          ||::|||| |||:|   |::   :|   :| |: |:  | |:  |  | :  ||| |
Qy    237 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQ-- 294

Db    337 svaqayysydpsskelisfdtpdmintkvsylknlglggsmfw 379
            : :   :|    :: ::||  : ::|||||||  ||||:| |
Qy    295 RIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT    8
ID   R56861 standard; Protein; 389 AA.
AC   R56861;
DT   17-FEB-1995  (first entry)
DE   Aphanocladium album mature endochitinase.
KW   oxalic acid degradation; oxalate oxidase; selection marker;
KW   filamentous fungus endochitinase; precursor;
KW   pathogen resistance; plant cell selection.
OS   Aphanocladium album.
FH   Key             Location/Qualifiers
FT   mat_peptide     1..1167
FT   /*tag=  a
FT   /product= endochitinase
PN   WO9413790-A.
PD   23-JUN-1994.
PF   07-DEC-1993; F01203.
PR   07-DEC-1992; FR-014721.
PA   (SNFI ) ELF SANOFI.
PA   (ERAP ) SOC NAT ELF AQUITAINE.
PI   Grezes-Besset B,  Grison R,  Pignard A,  Schneider M;
DR   WPI; 94-217870/26.
DR   N-PSDB; Q68353.
PT   Use of gene encoding oxalic acid degrading enzyme - for plant
PT   cell selection, esp. gene coupled to gene-encoding protein
PT   conferring pathogen resistance
PS   Disclosure and Claim 8; Page 39-40; 64pp; French.
CC   The sequence coding for wheat germin, a protein induced during
CC   germination of wheat, is a preferred example of DNA coding for a
CC   protein able to degrade oxalic acid. Such DNA can be coupled to a
CC   sequence of interest, e.g. one coding for enzymes which confer
CC   resistance to pathogens (esp. endochitinases). Plant calli which
CC   have been successfully transformed by the sequences are selected
CC   by their ability to grow on oxalic acid-contg. medium. R56861 is
CC   a known sequence of a mature endochitinase; the sequence was
CC   previously disclosed in EP-531218.
SQ   Sequence   389 AA;

  Query Match           21.2%;  Score 576;  DB 10;  Length 389;
  Best Local Similarity  36.2%;  Pred. No. 1.01e-41;
  Matches   101;  Conservative   65;  Mismatches 95;  Indels  18;  Gaps  16;

Db     81 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 139
          ||| | ::| :|:|||  :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    140 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 194
           :::: | ::  :   |:   :::  :  : |  ::||| | ||| : |    :  : ::
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 180

Db    195 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 253
          ||::|||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    254 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 305
          |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 300

Db    306 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 344
```

```
          | :    :: ::||    : |||| || |||||:| |
Qy    301 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT    9
ID   R33072 standard; Protein; 423 AA.
AC   R33072;
DT   30-JUN-1993  (first entry)
DE   Aphanocladium album pre-pro-endochitinase.
KW   filamentous fungus; pathogen resistance.
OS   Aphanocladium album.
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT   /label= signal
FT   /note= "pre-sequence"
FT   Peptide         23..34
FT   /note= "pro-sequence"
FT   Protein         35..423
FT   /label= endochitinase
PN   EP-531218-A.
PD   10-MAR-1993.
PF   04-SEP-1992; 402414.
PR   06-SEP-1991; FR-011072.
PA   (SNFI ) ELF SANOFI.
PA   (ERAP ) SOC NAT ELF AQUITAINE.
PI   Blaiseau P,  Legoux R,  Leguay J,  Schneider M;
DR   WPI; 93-078700/10.
DR   N-PSDB; Q37711, Q37712.
PT   Recombinant DNA encoding fungal endo:chitinase - useful for
PT   protecting plants against fungi, insects, etc.
PS   Claim 1; Fig 1; 75pp; French.
CC   Clone CH3C was isolated from a cDNA bank prepared from A.album
CC   mRNA by screening with anti-chitinase antiserum. Three probes were
CC   designed based on the CH3C sequence, corresponding to the start,
CC   middle and downstream regions of the coding sequence. The probes
CC   were used to isolate the full-length coding sequence of A.album
CC   endochitinase from an A.album cDNA library in pTZ19R. A 1.6kb
CC   fragment was identified in several positive clones. The sequence
CC   of the fragment contained an ORF encoding a 423 amino acid protein.
CC   The mature endochitinase can be used to control pathogens
CC   such as fungi, bacteria, arthropods and nematodes on plants.
SQ   Sequence   423 AA;

  Query Match           21.2%;  Score 576;  DB 6;  Length 423;
  Best Local Similarity  36.2%;  Pred. No. 1.01e-41;
  Matches   101;  Conservative   65;  Mismatches  95;  Indels  18;  Gaps  16;

Db    115 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 173
          ||| | ::| :|:|||| :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    174 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 228
           :::: | ::  :  |:   ::: :  : |  ::||| | ||| : |    :  : ::
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 180

Db    229 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 287
           ||::||||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    288 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 339
          |||| |||||   :  :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 300

Db    340 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 378
          | :    :: ::||    : |||| || |||||:| |
Qy    301 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT   10
ID   R33068 standard; Protein; 389 AA.
AC   R33068;
DT   30-JUN-1993  (first entry)
DE   Aphanocladium album mature endochitinase.
KW   filamentous fungus; pathogen resistance.
OS   Aphanocladium album.
PN   EP-531218-A.
PD   10-MAR-1993.
PF   04-SEP-1992; 402414.
PR   06-SEP-1991; FR-011072.
PA   (SNFI ) ELF SANOFI.
PA   (ERAP ) SOC NAT ELF AQUITAINE.
PI   Blaiseau P,  Legoux R,  Leguay J,  Schneider M;
DR   WPI; 93-078700/10.
DR   N-PSDB; Q37704.
PT   Recombinant DNA encoding fungal endo:chitinase - useful for
PT   protecting plants against fungi, insects, etc.
PS   Claim 22; Page 60-61; 75pp; French.
CC   This mature protein from the filamentous fungus Aphanocladium
CC   album has endochitinase activity and is used to control pathogens
CC   such as fungi, bacteria, arthropods and nematodes on plants.
CC   Recombinant DNA comprising the endochitinase coding sequence is
CC   specifically intended for transforming Nicotiana tabacum, Helianthus
CC   annuus and Brassica napus to pathogen resistance.
SQ   Sequence   389 AA;

  Query Match           20.9%;  Score 569;  DB 6;  Length 389;
  Best Local Similarity  35.8%;  Pred. No. 4.58e-41;
  Matches   100;  Conservative   66;  Mismatches  95;  Indels  18;  Gaps  16;

Db     81 lkkqnrnmkvmvsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 139
          ||| | ::| :::|||| :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    140 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 194
           :::: | ::  :  |:   ::: :  : |  ::||| | ||| : |    :  : ::
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 180

Db    195 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 253
           ||::||||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    254 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 305
          |||| |||||   :  : :| |  | |: |:   | |:  |  :    |||    :  |
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 300

Db    306 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 344
          | :    :: ::||    : |||| || |||||:| |
Qy    301 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT   11
ID   W02159 standard; Protein; 866 AA.
AC   W02159;
DT   14-JAN-1997  (first entry)
DE   Soluble chitinase.
KW   Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;
KW   Beta-N-acetylglucosamidase; chitin; oligosaccharide; catabolic;
KW   catabolism.
OS   Vibrio furnissii.
PN   WO9625424-A1.
PD   22-AUG-1996.
PF   13-FEB-1996; U02332.
PR   13-FEB-1995; US-386727.
PA   (UYJO ) UNIV JOHNS HOPKINS.
PI   Bassler B,  Chitlaru E,  Keyhani N,  Roseman S,  Rowe C;
PI   Yu C;
DR   WPI; 96-393335/39.
DR   N-PSDB; T36390.
PT   Chitin biosynthetic enzymes end I, exo I and exo II  - are
PT   periplasmic chito:dextrinase(s), periplasmic beta-GlcNAcidase(s) and
PT   aryl beta-N-acetyl:gluco:amidase(s), respectively
PS   Example 4; Page 79-82; 101pp; English.
CC   Periplasmic chitodextrinase (W02156), periplasmic
CC   Beta-N-acetylglucosaminidase (W02157) and aryl
CC   Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin
CC   oligosaccharides with the structure (GlcNAc)n where n is 2 or
```

```
CC   higher, by contacting them with soluble chitin.  The enzymes are
CC   encoded by the genes endI, exoI and exoII respectively.  They are
CC   all genes involved in the catabolic pathway of chitin.
SQ   Sequence   866 AA;

  Query Match           17.0%;  Score 463;  DB 19;  Length 866;
  Best Local Similarity  30.7%;  Pred. No. 3.52e-31;
  Matches    93;  Conservative   81;  Mismatches 104;  Indels  25;  Gaps  20;

Db    253 fgnlmalkkanpnlkilpsvggwtlsdp-fy-ffsdktkrdtfvasmkeylqtwkffdgv 310
          : :: :||| ||:|| | ::|||  ::    |    :   :|:||| |   :|:   | ||:
Qy     56 YQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSF-DGL 114

Db    311 didwefpggqganpnlggpndgatyvalmkelramldeleaet-grqyeltsaisaggdk 369
          |:|||:||:||: | :   ::   : :|: :|   ::: ||:| |::  | ||  ::|:
Qy    115 DLDWEYPGSQGS-PAVD--KE--RFTTLVQDLANAFQQ-EAQTSGKERLLLSAAVPAGQT 168

Db    370 iakvdyqaaq--qymdyiflmsydfsgafdlknlahqtnly--asswdpatkyttdkgvk 425
                |:    | :|:: ||:||| |: : |  :| : ||      :|:    | :|
Qy    169 YVDAGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNSPLYKRQEESGAAASLNVDAAVQ 227

Db    426 allgqgvtpgkvvvgaamygrgwt-gvng-yqagnpftgsat-gpikgtwengvvdyrdi 482
            |  |  ::|:::| : |||: | : ::  : | | |||:| ||:    | |:: | ::
Qy    228 QWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFT-K-EGGMLAYYEV 285

Db    483 vnnrmgagweqgydetaeapyvfkastgdlisfdndrsvkakgqyvlanqlgglfaweid 542
           : ::||   |  ::    ||:|: :  : ::||:  | |:|  |:  : |||   | :|
Qy    286 CS-WKGAT-KQRIQD-QKVPYIFRDN--QWVGFDDVESFKTKVSYLKQKGLGGAMVWALD 340

Db    543 adn 545
            |:
Qy    341 LDD 343


RESULT   12
ID   R64823 standard; Protein; 561 AA.
AC   R64823;
DT   23-JUL-1995  (first entry)
DE   Chitinase A.
KW   Chitinase A; protein secretion; biological control agent;
KW   transgenic plant; chitin; plant pathogen; nematode; fungus;
KW   crop protection; pCHIT1251.
OS   Serratia marcescens QMB1466.
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT   /label= Sig_peptide
PN   US5374540-A.
PD   20-DEC-1994.
PF   26-MAR-1984; 593691.
PR   26-MAR-1984; US-593691.
PR   18-JUL-1986; US-888033.
PR   09-JUL-1990; US-550253.
PA   (DNAP ) DNA PLANT TECHNOLOGY CORP.
PI   Jones JDG,  Suslow T;
DR   WPI; 95-035648/05.
DR   N-PSDB; Q76290.
PT   Plant cell contg. signal sequence of chitinase A gene - for
PT   secretion of foreign polypeptide, esp. chitinase A for protection
PT   against pathogenic fungi and nematodes.
PS   Disclosure; Fig. 1A-1C; 20pp; English.
CC   Serratia marcescens QMB1466 genomic DNA was partially digested,
CC   ligated into vector pLAFR1, packaged into lambda phage particles,
CC   and used to transfect Escherichia coli DH1 (ATCC 33849) cells.
CC   Plating on chitin medium was used to screen for chitinase activity.
CC   Clone C12 (ATCC 67152) was isolated.  Its sequence is given in
CC   Q76290, and the deduced protein sequence in R64823.
SQ   Sequence   561 AA;

  Query Match           14.3%;  Score 389;  DB 12;  Length 561;
  Best Local Similarity  28.2%;  Pred. No. 2.29e-24;
  Matches    91;  Conservative   81;  Mismatches 126;  Indels  25;  Gaps  18;

Db    231 pfaalqkaqkgvtawddpykgnfgqlmalkqahpdlkilpsiggwtlsdpfff-mgdkvk 289
          :||:: : | : | |:|     : :: :||  :| || | :|||| ::   |  |    :
Qy     36 AFAGMTNHQLSTTEWNDE-T-LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATAN 93

Db    290 -rdrfvgsvkeflqtwkffdgvdidwefpggkganpnlgspqdgetyvllmkelramldq 348
           |: || |   ||:   | ||:|:|||:||: |: | :    | | :  |: :|   ::|
Qy     94 NRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV----DKERFTTLVQDLANAFQQ 147

Db    349 lsaetgr-kyeltsaisagkdkidkvaynvaq--nsmdhiflmsydfygafdlknlghqt 405
          :  :|: :  |::|::||   :|  :|:|     ::| : ||:||| |: : |  || :
Qy    148 EAQTSGKERLLLSAAVPAGQTYVD-AGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNS 205

Db    406 alnar---pgsrhrlhhgerrecaagqgvkpgkivvgtamygrgwtgvngyqnnipftgt 462
          :|  |   :|:   |:      :     |  ::|:::| : |||: |  :: :  :   :|
Qy    206 PLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPAT 265

Db    463 hravkgtw--engivdyrqiasqfmsgewqytydataeapyvfkpstgdlitsddarsvq 520
            :  |    | |:: | :: |  :::  |   :     ||:|:    : :  ||  |
Qy    266 GSGTPGPFTKEGGMLAYYEVCS-WKGATKQ-RIQD-QKVPYIFR-DN-QWVGFDDVESFK 320

Db    521 akgkyvldkqlgglfsweidadn 543
          :|  |: :| |||   | :| |:
Qy    321 TKVSYLKQKGLGGAMVWALDLDD 343


RESULT   13
ID   R48981 standard; Protein; 561 AA.
AC   R48981;
DT   04-SEP-1994  (first entry)
DE   Bacterial chitinase-A fragment from pCHIT1251.
KW   Chitinase-A; enzyme; transgenic plant; biological control;
KW   disease-resistance; crop improvement; Agrobacterium; fungus;
KW   nematode; insect.
PN   US5290687-A.
PD   01-MAR-1994.
PF   26-MAR-1984; 593691.
PR   26-MAR-1984; US-593691.
PR   18-JUL-1986; US-888033.
PR   09-JUL-1990; US-550253.
PR   14-AUG-1992; US-930970.
PA   (DNAP ) DNA PLANT TECHNOLOGY CORP.
PI   Jones JDG,  Suslow T;
DR   WPI; 94-074348/09.
DR   N-PSDB; Q57966.
PT   Inhibition of chitinous fungal plant pathogens - by transforming
PT   plant with DNA encoding chitinase activity
PS   Disclosure; Fig.1; 13pp; English.
CC   Plasmid pCHIT1251 containing the chitinase-A gene (chiA), fused
CC   directly to a nos promoter,  may be used to transform, via
CC   Agrobacterium sp., a crop plant.  The resulting transgenic plants
CC   expressing the chitinase-A protein fragment are protected against
CC   fungi, nematodes, insects and disease agents. The protein is
CC   preferably expressed by the plant only at the site of pathogen
CC   attack, e.g. only in the root cells.  The plants express the
CC   chitinase in a biologically active form.
SQ   Sequence   561 AA;

  Query Match           13.7%;  Score 372;  DB 9;  Length 561;
  Best Local Similarity  27.6%;  Pred. No. 8.14e-23;
  Matches    89;  Conservative   84;  Mismatches 125;  Indels  25;  Gaps  18;

Db    231 pfaalqkaqkgvtawddpykgnfgqlmalkqahpdlkilpsiggwtlsdpfff-mgdkvk 289
          :||:: : | : | |:|     : :: :||  :| || | :|||| ::   |  |    :
Qy     36 AFAGMTNHQLSTTEWNDE-T-LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATAN 93

Db    290 -rdrfvgsvkeflqswkffdgvdidwefpggkganpnlgspqdgetyvlvmkevgamldq 348
           |: || |   ||:   | ||:|:|||:||: |: | :    | | :  :: :::  ::|
Qy     94 NRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV----DKERFTTLVQDLANAFQQ 147

Db    349 lwaetgr-kyeltsaisagkdkidkvaynvaq--nsmdhiflmsydfygafdlknlghqt 405
            :|: :  |::|::||   :|  :|:|     ::| : ||:||| |: : |  || :
Qy    148 EAQTSGKERLLLSAAVPAGQTYVD-AGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNS 205

Db    406 alnar---pgsrhrlhhgegrecaagqgvkpgkivvgtamygrgwtgvngyqnnipftgt 462
```

```
           :|  |   :|:   |:   : :     |  ::|:::| : |||: |  :: :  :   :|
Qy     206 PLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPAT 265

Db     463 hravkgtw--engivdyrqiasqfmsgewqytydgtaeapyvfkpstgdlitsddarsvq 520
           :  |    | |:: | :: |  :::  |   :     ||:|:    : :  ||  |
Qy     266 GSGTPGPFTKEGGMLAYYEVCS-WKGATKQ-RIQD-QKVPYIFR-DN-QWVGFDDVESFK 320

Db     521 akgkyvldkqlgglfsweidadn 543
           :|  |: :| |||   | :| |:
Qy     321 TKVSYLKQKGLGGAMVWALDLDD 343


RESULT   14
ID   R32547 standard; Protein; 238 AA.
AC   R32547;
DT   17-JUN-1993  (first entry)
DE   HaNPV IE-1 gene protein (partial sequence).
KW   Heliothis armigera nuclear polyhedrosis virus; Ha; NPV;
KW   insecticide; immediate early-1.
OS   Heliothis armigera nuclear polyhedrosis virus.
PN   WO9303144-A.
PD   18-FEB-1993.
PF   05-AUG-1992; AU0413.
PR   05-AUG-1991; AU-007576.
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PI   Christian PD, Dall DJ, Gordon KHJ, Hanzlik TN, Sriskantha A.
DR   WPI; 93-076501/09.
DR   N-PSDB; Q35982.
PT   Insect virus with reduced capacity to occlude viral particles - used
PT   for controlling proliferation of insect pests without horizontal
PT   transmission
PS   Disclosure; Fig 2; 51pp; English.
CC   The sequence is that encoded by the Heliothis armigera (Ha)
CC   nuclear polyhedrosis virus (NPV) immediate early-1 (IE-1) gene
CC   (partial sequence).
SQ   Sequence   238 AA;

  Query Match           10.0%;  Score 271;  DB 6;  Length 238;
  Best Local Similarity  34.1%;  Pred. No. 9.44e-14;
  Matches    58;  Conservative   43;  Mismatches  51;  Indels  18;  Gaps  14;

Db      44 erdkflsnpgwrefvltwkf-fdgididwefpggkganpnvgdverdnntyiallgelra 102
           :|: |::::  | | | :|: |||:|:|||:||: |: | | | ||    : :|: :|
Qy      94 NRQTFVNSAI-R-F-L-RKYSFDGLDLDWEYPGSQGS-PAV-DKER----FTTLVQDLAN 143

Db     103 mldq-vqiqtnrtlelttaisagidkiaav-ncdraqqyldkifvmsydfkgawsntdlg 160
            ::|  |   :  | |::|::||   : |  : |:  | || : :|:||| |:| :   |
Qy     144 AFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVT-G 202

Db     161 hqtaly--gsawkpnepytanvavdallaqrvnp-kklvlgvamygrgwt 207
           | ::||       :  :   : ||:  | |:  |  ||:||:: |||: |
Qy     203 HNSPLYKRQEESGAAASLNVDAAVQQWL-QKGTPASKLILGMPTYGRSFT 251


RESULT   15
ID   W02156 standard; Protein; 1046 AA.
AC   W02156;
DT   14-JAN-1997  (first entry)
DE   Periplasmic chitodextrinase.
KW   Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;
KW   Beta-N-acetylglucosamidase; chitin; oligosaccharide; catabolic;
KW   catabolism.
OS   Vibrio furnissii.
PN   WO9625424-A1.
PD   22-AUG-1996.
PF   13-FEB-1996; U02332.
PR   13-FEB-1995; US-386727.
PA   (UYJO ) UNIV JOHNS HOPKINS.
PI   Bassler B,  Chitlaru E,  Keyhani N,  Roseman S,  Rowe C;
PI   Yu C;
DR   WPI; 96-393335/39.
DR   N-PSDB; T36387.
PT   Chitin biosynthetic enzymes end I, exo I and exo II  - are
PT   periplasmic chito:dextrinase(s), periplasmic beta-GlcNAcidase(s) and
PT   aryl beta-N-acetyl:gluco:amidase(s), respectively
PS   Claim 2; Page 68-71; 101pp; English.
CC   Periplasmic chitodextrinase (W02156), periplasmic
CC   Beta-N-acetylglucosaminidase (W02157) and aryl
CC   Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin
CC   oligosaccharides with the structure (GlcNAc)n where n is 2 or
CC   higher, by contacting them with soluble chitin.  The enzymes are
CC   encoded by the genes endI, exoI and exoII respectively.  They are
CC   all genes involved in the catabolic pathway of chitin.
SQ   Sequence   1046 AA;

  Query Match           8.1%;  Score 219;  DB 19;  Length 1046;
  Best Local Similarity  33.3%;  Pred. No. 3.13e-09;
  Matches    48;  Conservative   40;  Mismatches  42;  Indels  14;  Gaps  10;

Db     454 etfadsavemmrkyrfdgldidlrisnidggtgnpddtafsesrraylmnsyhelmrvlr 513
           :|| :||:  :||| ||||||:| :  :  |: |:|      | | : |:   ::|   ::
Qy      96 QTFVNSAIRFLRKYSFDGLDLD-W--EYPGSQGSPA-VD-KE-RFTTLV---QDLANAFQ 146

Db     514 ekldvasaqdgvhymltiaapss-ayllrgmetmavtqyldyvnimsydlhgawndhvgh 572
           :  : :|: :  : :|: | |:: :|:  | |   ::| ||:||:|:||:||:|:   ||
Qy     147 QEAQ-TSGKE--RLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGH 203

Db     573 naalydtgkdselaqw-nvygtaq 595
           |::||    :|  |   || :: |
Qy     204 NSPLYKRQEESGAAASLNVDAAVQ 227


Search completed: Fri Jun 27 17:10:12 1997
Job time : 102 secs.
```

****

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 27 17:04:42 1997; MasPar time 17.76 Seconds
598.666 Million cell updates/sec

Tabular output not generated.

Title: >US-08-663-618A-15
Description: (1-373) from US08663618A.pep
Perfect Score: 2717
Sequence: 1 AKLVCYFTNWAQYRQGEARF.....YPLIQTLRQELSLPYLSSGT 373

Scoring table: PAM 150
Gap 11

Searched: 89912 seqs, 28507787 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir50
1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8
13:unann9 14:unann10 15:unenc 16:unrev

Statistics: Mean 47.342; Variance 104.126; scale 0.455

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1488	54.8	383	13	A49562	cartilage glycoprote	1.46e-250
2	1428	52.6	383	16	S51327	heparin-binding glyc	4.41e-239
3	1417	52.2	381	14	S61550	BRP39 protein - mous	5.60e-237
4	1417	52.2	381	14	I48271	BRP39 protein - mous	5.60e-237
5	1345	49.5	399	5	S27879	secretory protein YM	3.17e-223
6	1328	48.9	537	14	S57197	oviduct-specific gly	5.55e-220
7	1318	48.5	539	14	I46470	estrogen dependent o	4.49e-218
8	1179	43.4	483	12	A53918	chitinase (EC 3.2.1.	1.28e-191
9	1086	40.0	504	12	A38221	chitinase (EC 3.2.1.	5.49e-174
10	1015	37.4	554	12	A56596	chitinase (EC 3.2.1.	1.44e-160
11	698	25.7	699	10	A38368	chitinase (EC 3.2.1.	2.65e-101
12	594	21.9	378	5	S51591	chitinase (EC 3.2.1.	3.36e-82
13	582	21.4	424	5	S47133	chitinase (EC 3.2.1.	5.18e-80
14	576	21.2	423	5	JQ1975	chitinase (EC 3.2.1.	6.42e-79
15	561	20.6	423	12	S51369	chitinase - fungus (	3.43e-76
16	546	20.1	427	12	JC4565	chitinase (EC 3.2.1.	1.80e-73
17	446	16.4	563	16	S60651	chitinase precursor	1.61e-55
18	404	14.9	561	9	A25090	chitinase (EC 3.2.1.	4.21e-48
19	344	12.7	499	9	S52422	chitinase (EC 3.2.1.	1.12e-37
20	345	12.7	799	8	PC4106	chitinase (EC 3.2.1.	7.52e-38
21	344	12.7	820	16	A40633	chitinase (EC 3.2.1.	1.12e-37
22	337	12.4	499	9	S04856	chitinase (EC 3.2.1.	1.77e-36
23	314	11.6	511	11	S61166	probable membrane pr	1.46e-32
24	287	10.6	610	10	JH0573	chitinase (EC 3.2.1.	5.03e-28
25	266	9.8	452	12	JC4038	47K glycoprotein - f	1.50e-24
26	263	9.7	1146	11	S07915	RF2 protein - yeast	4.67e-24
27	246	9.1	413	5	JC2135	chitinase (EC 3.2.1.	2.73e-21
28	229	8.4	124	5	S57715	chitinase (EC 3.2.1.	1.45e-18
29	172	6.3	36	14	A27682	39K whey protein - b	7.97e-10
30	172	6.3	83	13	A37954	estradiol-stimulated	7.97e-10
31	136	5.0	29	14	S57204	oviduct-specific gly	9.06e-05
32	136	5.0	29	14	S55693	oviduct-specific sia	9.06e-05
33	121	4.5	597	5	S33848	chitinase (EC 3.2.1.	8.00e-03
34	121	4.5	597	5	S32039	chitinase (EC 3.2.1.	8.00e-03
35	112	4.1	642	16	S37623	homologies with orf	1.02e-01
36	110	4.0	488	10	A41961	chitinase (EC 3.2.1.	1.78e-01
37	104	3.8	305	9	S28440	probable coproporphy	8.92e-01
38	104	3.8	422	4	ZWECC	umuC protein - Esche	8.92e-01
39	100	3.7	364	11	S45925	hypothetical protein	2.53e+00
40	101	3.7	562	3	KRHUEB	keratin, 56K type II	1.95e+00
41	101	3.7	662	11	S61193	probable membrane pr	1.95e+00
42	101	3.7	1045	2	SUSEMM	serine proteinase (E	1.95e+00
43	101	3.7	1070	11	S19686	alpha-glucosidase (E	1.95e+00
44	99	3.6	398	5	A37768	streptococcal pyroge	3.26e+00
45	99	3.6	958	1	JN0102	glucan 1,4-alpha-glu	3.26e+00

ALIGNMENTS

```
RESULT    1
ENTRY             A49562     #type complete
TITLE             cartilage glycoprotein gp39 precursor - human
ALTERNATE_NAMES   39K synovial protein
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change
                    03-May-1996
ACCESSIONS        A49562; S10677; A33162
REFERENCE         A49562
   #authors       Hakala, B.E.; White, C.; Recklies, A.D.
   #journal       J. Biol. Chem. (1993) 268:25803-25810
   #title         Human cartilage gp-39, a major secretory product of articular
                    chondrocytes and synovial cells, is a mammalian member of a
                    chitinase protein family.
   #accession     A49562
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-383 ##label HAK
      ##cross-references GB:M80927
REFERENCE         S10677
   #authors       Nyirkos, P.; Golds, E.E.
   #journal       Biochem. J. (1990) 269:265-268
   #title         Human synovial cells secrete a 39 kDa protein similar to a
                    bovine mammary protein expressed during the non-lactating
                    period.
   #cross-references MUID:90328983
   #accession     S10677
      ##molecule_type protein
      ##residues       22-40,'X',42-45 ##label NY2
KEYWORDS          cartilage; extracellular protein; glycoprotein
FEATURE
   1-21                #domain signal sequence #status predicted #label SIG\
   22-383              #product cartilage glycoprotein gp39 #status predicted
                         #label MAT
SUMMARY           #length 383  #molecular-weight 42613  #checksum 2942

  Query Match           54.8%;  Score 1488;  DB 13;  Length 383;
  Best Local Similarity  53.3%;  Pred. No. 1.46e-250;
  Matches   195;  Conservative   83;  Mismatches  79;  Indels   9;  Gaps   5;

Db     23 klvcyytswsqyregdgscfpdaldrflcthiiysfanisndhidtwewndvtlygmlnt 82
          |||||:|:|:|||:|::  :|  ||  ||||:||:|| ::| :: | |||| |||  :|
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDETLYQEFNG 61

Db     83 lknrnpnlktllsvggwnfgsqrfskiasntqsrrtfiksvppflrthgfdgldlawlyp 142
```

```
          ||: ||:|||||::||||||:|:|: : : : :|:||::|   |||  :||||||  | ||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    143 grr-----dkqhfttlikemkaefikeaq-pgkkqlllsaalsagkvtidssydiakisq 196
          | :     ||::|||: ::   |  ||| :|| :|||||::||   :|:::|:: ||:|
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQ 181

Db    197 hldfisimtydfhgawrgttghhsplfrgqedaspdrfsntdyavgymlrlgapasklvm 256
          :|||:::|:|||||:|   |||:|||:: ||::::    | | ||   |: |:|||||::
Qy    182 NLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLIL 241

Db    257 giptfgrsftlass-etgvgapisgpgipgrftkeagtlayyeicdflrgatvhrtlgqq 315
          |:||:|||||||| :| |||| :|:| || ||||:| |||||:|   :||| :|   |
Qy    242 GMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQK 300

Db    316 vpyatkgnqwvgyddqesvkskvqylkdrqlagamvwaldlddfqgsfcgqdlrfpltna 375
          |||  : |||||:|| || |:|| |||:: |:|||||||||||| |  | |  |:||  :
Qy    301 VPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQG-RYPLIQT 359

Db    376 ikdala 381
          ::: |:
Qy    360 LRQELS 365


RESULT    2
ENTRY             S51327     #type complete
TITLE             heparin-binding glycoprotein 38K - pig
ORGANISM          #formal_name Sus scrofa domestica #common_name domestic pig
DATE              07-May-1995 #sequence_revision 07-May-1995 #text_change
                    07-May-1995
ACCESSIONS        S51327
REFERENCE         S51327
   #authors       Shackelton, L.M.; Mann, D.M.; Millis, A.J.T.
   #submission    submitted to the EMBL Data Library, January 1995
   #description   Identification of a 38kDa heparin-binding glycoprotein
                    (gp38k) in differentiating vascular smooth muscle cells as
                    a member of a group of proteins associated with tissue
                    remodeling.
   #accession     S51327
      ##status         preliminary
      ##residues       1-383 ##label SHA
      ##cross-references EMBL:Z47803
SUMMARY           #length 383  #molecular-weight 42443  #checksum 2907

  Query Match           52.6%;  Score 1428;  DB 16;  Length 383;
  Best Local Similarity  52.2%;  Pred. No. 4.41e-239;
  Matches   191;  Conservative   81;  Mismatches  85;  Indels   9;  Gaps   5;

Db     23 klvcyytswsqyregdgscfpdaidpflcthiiysfanisnneidtlewndvtlydtlnt 82
          |||||:|:|:|||:|::  :|  :|| ||||:||:|| ::|::: | |||| |||: :|
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDETLYQEFNG 61

Db     83 lknrnpnlktllsvggwnfgsqrfskiasntqsrrtfiksvppflrthgfdgldlawisp 142
          ||: ||:|||||::||||||:|:|: : : : :|:||::|   |||  :||||||  |  |
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    143 grr-----dkrhlttlvkemkaefvreal-pgterlllsgavsagkvaidrgydiaqisq 196
          | :     || ::|||| ::   | :||  :| ||||||:||:||   :| ||::  |:|
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQ 181

Db    197 hldfislltydfhgawrqttghhsplfrgqgdassdrfsnadyavsyvlrlgapanklvm 256
          :|||:::|::||||||:|   |||:|||:: | ::::    | | ||   |: |:|||:||::
Qy    182 NLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLIL 241

Db    257 giptfgrsftlassk-tdvgapasgpgipgrftkekgilayyeicdflqgatvrrplgqq 315
          |:||:|||||||||  | |||||:|:| || |||| |:||||||:|    ||| :|   |
Qy    242 GMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQK 300

Db    316 vpyatkgnqwvgyddqesvknkakylksrqlagamvwtldlddfrgnfcgqnlrfpltsa 375
          |||  : |||||:|| || | |  ||| : |:||||||:|||||| |  | |  |:||  :
Qy    301 VPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQG-RYPLIQT 359

Db    376 ikdvla 381
          ::: |:
Qy    360 LRQELS 365


RESULT    3
ENTRY             S61550     #type complete
TITLE             BRP39 protein - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change
                    03-May-1996
ACCESSIONS        S61550
REFERENCE         S61550
   #authors       Morrison, B.W.
   #submission    submitted to the EMBL Data Library, November 1995
   #accession     S61550
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-381 ##label MOR
      ##cross-references EMBL:X93035
SUMMARY           #length 381  #molecular-weight 43001  #checksum 5107

  Query Match           52.2%;  Score 1417;  DB 14;  Length 381;
  Best Local Similarity  50.7%;  Pred. No. 5.60e-237;
  Matches   186;  Conservative   84;  Mismatches  86;  Indels  11;  Gaps   9;

Db     23 klvcyftswsqyregvgsflpdaiqpflcthiiysfanissdnmlstwewndesnydkln 82
          |||||||:|:||||: : |||  ::| ||||:||:|| :: :: ||| |||||: |: :|
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMT-NHQLSTTEWNDETLYQEFN 60

Db     83 klktrntnlktllsvggwkfgekrfseiasnterrtafvrsvapflrsygfdgldlawly 142
           ||  | :|||||::|||:||  :|::: : :: | :|| |   ||| |:|||||| | |
Qy     61 GLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEY 120

Db    143 p--r-lr--dkqyfstlikelnaeftkevq-pgreklllsaalsagkvaidtgydiaqia 196
          |  :     ||: |:||: :|   |  | | :|:|:|||||::||   :|:||::  ||
Qy    121 PGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIA 180

Db    197 qhldfinlmtydfhgvwrqitghhsplfqgqkdtrfdrysnvnyavqymirlgaqaskll 256
          |:|||:|||:|||||  |  :|||:||||:   | ::      ||: |||  :: |: ||||:
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    257 mgiptfgksftlass-enqlgapisgeglpgrftkeagtlayyeicdflkgaevhrlsne 315
          :|:||:|:|||||||  : ::||| :| | || ||||:| |||||:|   |||  :|: ::
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQ 299

Db    316 kvpfatkgnqwvgyehkesvknkvgflkekklagamvwaldlddfqg-tcqpkeffpltn 374
          |||:  : |||||::  || | ||::||:| |:||||||||||| | :|    : ||
Qy    300 KVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRY-PLIQ 358

Db    375 aikdala 381
          :::: |:
Qy    359 TLRQELS 365


RESULT    4
ENTRY             I48271     #type complete
TITLE             BRP39 protein - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
                    02-Jul-1996
ACCESSIONS        I48271
REFERENCE         I48271
   #authors       Morrison, B.W.; Leder, P.
   #journal       Oncogene (1994) 9:3417-3426
   #title         neu and ras initiate murine mammary tumors that share genetic
                    markers generally absent in c-myc and int-2-initiated
                    tumors.
   #cross-references MUID:95060797
   #accession     I48271
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-381 ##label RES
      ##cross-references EMBL:X93035; NID:g1085065; CDS_PID:g1085066
```

```
GENETICS
   #note          gene name brp39
SUMMARY           #length 381  #molecular-weight 43001  #checksum 5107

  Query Match           52.2%;  Score 1417;  DB 14;  Length 381;
  Best Local Similarity  50.7%;  Pred. No. 5.60e-237;
  Matches   186;  Conservative   84;  Mismatches  86;  Indels  11;  Gaps   9;

Db      23 klvcyftswsqyregvgsflpdaiqpflcthiiysfanissdnmlstwewndesnydkln 82
           |||||||:|:|||:| : |||  ::| ||||:||:|| :: :: ||| |||||: |: :|
Qy       2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMT-NHQLSTTEWNDETLYQEFN 60

Db      83 klktrntnlktllsvggwkfgekrfseiasnterrtafvrsvapflrsygfdgldlawly 142
            ||  | :|||||::|||:||  :|::: : :: | :|| |   ||| |:|||||| | |
Qy      61 GLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEY 120

Db     143 p--r-lr--dkqyfstlikelnaeftkevq-pgreklllsaalsagkvaidtgydiaqia 196
           |  :     ||: |:||: :|   |  | | :|:|:||||||::||   :|:||::  ||
Qy     121 PGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIA 180

Db     197 qhldfinlmtydfhgvwrqitghhsplfqgqkdtrfdrysnvnyavqymirlgaqaskll 256
           |:|||:|||:|||||  |  :|||:|||:  | ::      ||: |||  :: |: ||||:
Qy     181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db     257 mgiptfgksftlass-enqlgapisgeglpgrftkeagtlayyeicdflkgaevhrlsne 315
           :|:||:|:|||||| : ::||| :| | || ||||:| |||||:|   |||  :|: ::
Qy     241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQ 299

Db     316 kvpfatkgnqwvgyehkesvknkvgflkekklagamvwaldlddfqg-tcqpkeffpltn 374
           |||:  : |||||::  || | ||::||:| |:||||||||||| | :|    : ||
Qy     300 KVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRY-PLIQ 358

Db     375 aikdala 381
           :::: |:
Qy     359 TLRQELS 365


RESULT    5
ENTRY             S27879     #type complete
TITLE             secretory protein YM-1 precursor - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
                    03-May-1996
ACCESSIONS        S27879
REFERENCE         S27879
   #authors       Chang, N.C.A.; Liu, C.H.; Chang, A.C.
   #submission    submitted to the EMBL Data Library, June 1992
   #description   Molecular characterization of a secretory protein (YM-1)
                    transiently expressed by activated murine peritoneal
                    macrophages.
   #accession     S27879
      ##molecule_type mRNA
      ##residues       1-399 ##label CHA
      ##cross-references EMBL:M94584
CLASSIFICATION    #superfamily Streptomyces chitinase chi40
FEATURE
   1-21               #domain signal sequence #status predicted #label SIG\
   22-399             #product secretory protein YM-1 #status predicted #label
                        MAT
SUMMARY           #length 399  #molecular-weight 44528  #checksum 3208

  Query Match           49.5%;  Score 1345;  DB 5;  Length 399;
  Best Local Similarity  47.7%;  Pred. No. 3.17e-223;
  Matches   175;  Conservative   87;  Mismatches 102;  Indels   3;  Gaps   3;

Db      24 lmcyytswakdrpiegsfkpgnidpclcthliyafagmqnneitytheqdlrdyealngl 83
           |:||:|:||  |  |: | | ::|| |||||||||||| |:::: |   |   |: :|||
Qy       3 LVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDETLYQEFNGL 62

Db      84 kdkkntelktllaiggwkfgpapfsamvstpqnrqifiqsvirflrqynfdglnldwqyp 143
           | |:|  ||||||||||:||   |: ||:|: ||| |: | ||||| |:||||:|||:||
Qy      63 K-KMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db     144 gsrgsppkdkhlfsvlvkemrkafeeesvekdiprllltstgagiidviksgtrslncls 203
           ||:|||: ||  |: || :: :||::|:      ||||::: ::    : :|    :
Qy     122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQ 181

Db     204 lldyiqvmtydlhdpkdgytgensplykspydigksadlnvdsiisywkdhgaasekliv 263
           ||:: :|:||:| : :  || ||||||   : | :| ||||: :  | : |::: |||:
Qy     182 NLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLIL 241

Db     264 gfpayghtfilsdpsktgigaptistgppgkytdesgllayyevctflnegatevwdapq 323
           | |:||::| |: :| | :|||: ::| || :| |:|:||||||:  : |||  :   |
Qy     242 GMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-K-GATKQRIQDQ 299

Db     324 evpyayqgnewvgydnvrsfklkaqwlkdnnlggavvwpldmddfsgsfchqrhfpltst 383
           ||| :: |:|||:|:| ||| |   ||:: ||||:||:||:|||:|  |:| ::||  |
Qy     300 KVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYPLIQT 359

Db     384 lkgdlni 390
           |: :|::
Qy     360 LRQELSL 366


RESULT    6
ENTRY             S57197     #type fragment
TITLE             oviduct-specific glycoprotein 95K precursor - bovine
                    (fragment)
ORGANISM          #formal_name Bos primigenius taurus #common_name cattle
DATE              28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
                    03-May-1996
ACCESSIONS        S57197
REFERENCE         S57197
   #authors       Sendai, Y.; Abe, H.; Kikuchi, M.; Satoh, T.; Hoshi, H.
   #journal       Biol. Reprod. (1994) 50:927-934
   #title         Purification and molecular cloning of bovine oviduct-specific
                    glycoprotein.
   #accession     S57197
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-537 ##label SEN
      ##cross-references EMBL:D16639
SUMMARY           #length 537  #checksum 1695

  Query Match           48.9%;  Score 1328;  DB 14;  Length 537;
  Best Local Similarity  51.2%;  Pred. No. 5.55e-220;
  Matches   187;  Conservative   74;  Mismatches  97;  Indels   7;  Gaps   6;

Db      20 klvcyftnwafsrpgpasilprdldpflcthlvfafasmsnnqivpkdpqdekilypefn 79
           ||||||||||  | | | :||:|||| |||||::|||:|:|:|:   :  ||  || |||
Qy       2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDET-LYQEFN 60

Db      80 klkernrglktllsiggwnfgtvrfttmlstfsnrerfvssviallrthgfdgldlffly 139
            ||  |  |||||:|||||||| :|| |::| :||: ||:| | :||  :||||||   |
Qy      61 GLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEY 120

Db     140 pglrgspardrwtfvflleellqafkneaqltmrprlllsaavsgdphvvqkayearllg 199
           || :|||| |:  |  |:::|  ||  ||| : : ||||||||::    |: :||   ::
Qy     121 PGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIA 180

Db     200 rlldfisvlsydlhgswekvtghnsplfs-l--pgdpks-sa-yamnywrqlgvppekll 254
           : |||:::::||:|||||||||||||||:     :| : | :   |:  | | | |: ||:
Qy     181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db     255 mglptygrtfhllkasqnelraqavgpaspgkytkqagflayyeiccfvrrakkrwindq 314
           :|:||||||:| |  :|:  : | | |:::|| :||::| |||||:|   : | |::| ||
Qy     241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQ 299

Db     315 yvpyafkgkewvgyddaisfgykaffikrehfggamvwtldlddfrgyfcgtgpfplvht 374
           ||| |: ::|||:|| ||  |  ::|:  :||||||:||||| |: |  | :||::|
Qy     300 KVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYPLIQT 359

Db     375 lnnll 379
           |   |
Qy     360 LRQEL 364
```

```
RESULT    7
ENTRY             I46470     #type complete
TITLE             estrogen dependent oviduct protein precursor - sheep
ORGANISM          #formal_name Ovis orientalis aries, Ovis ammon aries
                    #common_name domestic sheep
DATE              16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change
                    16-Aug-1996
ACCESSIONS        I46470
REFERENCE         I46470
   #authors       DeSouza, M.M.; Murray, M.K.
   #journal       Endocrinology (1995) 136:2485-2496
   #title         An estrogen-dependent secretory protein, which shares
                    identity with chitinases, is expressed in a temporally and
                    regionally specific manner in the sheep oviduct at the time
                    of fertilization and embryo development.
   #cross-references MUID:95269691
   #accession     I46470
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-539 ##label DES
      ##cross-references EMBL:U16719; NID:g885600; CDS_PID:g885601
SUMMARY           #length 539  #molecular-weight 59535  #checksum 2120

  Query Match           48.5%;  Score 1318;  DB 14;  Length 539;
  Best Local Similarity  50.1%;  Pred. No. 4.49e-218;
  Matches   183;  Conservative   77;  Mismatches  98;  Indels   7;  Gaps   5;

Db     23 klvcyftnwafsrpgsasilprdldpflcthlvfafasmnnnqivpkdpldekilypefn 82
        ||||||||||  | | | :||:|||| |||||::|||:| |:|:   :  ::  || |||
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTE-WNDETLYQEFN 60

Db     83 klkernrglktllsvggwnfgtsrftkmlstfsnrerfvksviallrthgfdgldlffly 142
          ||  |  |||||::|||||| :|| |::| :||: ||:| | :||  :||||||   |
Qy     61 GLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEY 120

Db    143 pglrgspardrwtfvflleellqafkneaqltmrprlllsaavsgdphviqkaydarllg 202
          || :|||| |:  |  |:::|  ||  ||| : : |||||||::    :: :|:   ::
Qy    121 PGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIA 180

Db    203 rlldfisvlsydlhgswekvtghnsplfslpgdpkssa-yam--sy--wrqlgvppekll 257
          : |||:::::||:|||||||||||||:    :: ::|   :  :   | | | |: ||:
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    258 mglptygrtfhllrasqnelgagaagpaspgkytkqagflayyevcsfvqrakkrwindq 317
          :|:|||||:| |  :|:  :|| |:|:::|| :||::| |||||||    | |::| ||
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQ 299

Db    318 yvpyafkgkewvgyddaisfgykaffikrehfggamvwtldlddfrgnfcgtgpfplaht 377
          ||| |: ::|||:||  ||  |  ::|:  :||||||:|||||| |  |  | :|| :|
Qy    300 KVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYPLIQT 359

Db    378 lnnll 382
          |   |
Qy    360 LRQEL 364


RESULT    8
ENTRY             A53918     #type complete
TITLE             chitinase (EC 3.2.1.14) precursor - braconid wasp (Chelonus
                    sp.)
ORGANISM          #formal_name Chelonus sp.
DATE              28-Jul-1995 #sequence_revision 28-Jul-1995 #text_change
                    03-May-1996
ACCESSIONS        A53918
REFERENCE         A53918
   #authors       Krishnan, A.; Nair, P.N.; Jones, D.
   #journal       J. Biol. Chem. (1994) 269:20971-20976
   #title         Isolation, cloning, and characterization of new chitinase
                    stored in active form in chitin-lined venom reservoir.
   #accession     A53918
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-483 ##label KRI
      ##cross-references GB:U10422
KEYWORDS          glycosidase; hydrolase
SUMMARY           #length 483  #molecular-weight 52013  #checksum 5938

  Query Match           43.4%;  Score 1179;  DB 12;  Length 483;
  Best Local Similarity  42.5%;  Pred. No. 1.28e-191;
  Matches   158;  Conservative   98;  Mismatches  99;  Indels  17;  Gaps  11;

Db     23 kvvcyfgawsvyrqgngkfdingidptlcthliysfvgvngkdvkvldpwsdlpgnldgf 82
          |:|||  |: ||||:::|  : :||:|||||||:| |:   ::   : |:|    |  :
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTE-WND--ETL--Y 56

Db     83 gkftslrkknpsvkimvavggwnagsvpfsqmasdqatreafaqnvvkflqqyqfdgfdi 142
            | :|:|:|| :| ::|:|||| |:  |::| :    |::|  : ::||: | |||:|:
Qy     57 QEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 116

Db    143 dweypaqrggspadvknmvklckalkkafvq--h----d-yilsaavaapetsasksydi 195
          |||||: :|::: |      |   | :|| |  :    :  :|||||:| :|    :|::
Qy    117 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 176

Db    196 aemsqyldfinlmtydfhgpwdghtgmhappsasshdsgnelklnvkaavkywlqngvpk 255
            ::| |||:|||:|||||:|:  || ::|      :||    ||| |||  |||:| |
Qy    177 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 236

Db    256 eklvvgvpaygksftlsnpsnkglgapvsgagtagpytgengllgyneicemqkagdwev 315
            ||::|:|:||:||||:::|:  :||| :|:||:||:| | |:|:| |:|   |::  :
Qy    237 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQR 295

Db    316 vqdnekgvpyavkgnqwvsfddlaaikakaqfikqeglggamvwsietddfkgl-cge-k 373
          :|| :| |||  : ||||:|||: ::|:|  ::|| ||||||||::: ||| |: | : :
Qy    296 IQD-QK-VPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGR 353

Db    374 ypvlkalnsvlg 385
          ||:: :|   |:
Qy    354 YPLIQTLRQELS 365


RESULT    9
ENTRY             A38221     #type complete
TITLE             chitinase (EC 3.2.1.14) MF1 - nematode (Brugia malayi)
ORGANISM          #formal_name Brugia malayi
DATE              04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
                    03-May-1996
ACCESSIONS        A38221
REFERENCE         A38221
   #authors       Fuhrman, J.A.; Lane, W.S.; Smith, R.F.; Piessens, W.F.;
                    Perler, F.B.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1992) 89:1548-1552
   #title         Transmission-blocking antibodies recognize microfilarial
                    chitinase in brugian lymphatic filariasis.
   #cross-references MUID:92179220
   #accession     A38221
      ##status         preliminary; not compared with conceptual translation
      ##molecule_type nucleic acid; protein
      ##residues       1-504 ##label FUH
      ##cross-references NCBIP:85345
      ##note           sequence extracted from NCBI backbone
KEYWORDS          glycosidase; hydrolase
SUMMARY           #length 504  #molecular-weight 55971  #checksum 7343

  Query Match           40.0%;  Score 1086;  DB 12;  Length 504;
  Best Local Similarity  42.8%;  Pred. No. 5.49e-174;
  Matches   158;  Conservative   94;  Mismatches 102;  Indels  15;  Gaps  11;

Db     27 cyytnwaqyrdgegkflpgnipnglcthilyafakvdelgdskpfewndedtewskgmys 86
          ||:|||||||:||::||| ::  :||||::|||| : :   |   |||||  |     :|
Qy      5 CYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTT-EWNDE-T-----LYQ 57

Db     87 avtklretnpglkvllsyggynfgsaiftgiaksaqkterfiksaiaflrknnfdgfdld 146
              |:  || || ||: || |||:  || :  :| : : |::||| |||| :|||:|||
Qy     58 EFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 117
```

```
Db    147 weyp-v-gv-a-e-eh-aklveamktafveeaktsgkqrllltaavsagkgtidgsynve 200
          ||||   |  | : |: : ||: :  || :|| ||||:||||:|||:||   :|::|:|:
Qy    118 WEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVD 177

Db    201 slgknfdllflmsydlhgsweknvdlhgklhptkgevsgigifntefaadywaskgmpke 260
           :: |:|:: ||:||:|||||     :: |     | :: : :| : | : |  || |
Qy    178 KIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPAS 237

Db    261 kiiigipmyaqgwtldnpsetaigaaasrpssasktnpaggtasyweickylkeggketv 320
          |:|:|:| |::: || ::|:|: :||:|: :::::     ||  :| |:|   | : |: :
Qy    238 KLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRI 296

Db    321 hqegvgaymvkgdqwygydneetirikmkwlkekgyggafiwaldfddftgkscgkgpyp 380
          ::: | :|: : :|| |:|: |::: |:  ||:|| ||| :||||:|||:| ||  | ||
Qy    297 QDQKV-PYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYP 355

Db    381 llnaissel 389
          |: ::  ||
Qy    356 LIQTLRQEL 364


RESULT   10
ENTRY            A56596     #type complete
TITLE            chitinase (EC 3.2.1.14) - tobacco hornworm
ORGANISM         #formal_name Manduca sexta #common_name tobacco hornworm
DATE             11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change
                   03-May-1996
ACCESSIONS       A56596
REFERENCE        A56596
   #authors      Kramer, K.J.; Corpuz, L.; Choi, H.K.; Muthukrishnan, S.
   #journal      Insect Biochem. Mol. Biol. (1993) 23:691-701
   #title        Sequence of a cDNA and expression of the gene encoding
                   epidermal and gut chitinases of Manduca sexta.
   #cross-references MUID:93357793
   #accession    A56596
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-554 ##label KRA
      ##cross-references NCBIN:136417; NCBIP:136418
      ##experimental_source larvae
      ##note          sequence extracted from NCBI backbone
KEYWORDS         glycosidase; hydrolase
SUMMARY          #length 554  #molecular-weight 62203  #checksum 4659

  Query Match           37.4%;  Score 1015;  DB 12;  Length 554;
  Best Local Similarity  40.4%;  Pred. No. 1.44e-160;
  Matches   155;  Conservative   98;  Mismatches 107;  Indels  24;  Gaps  18;

Db     23 arivcyfsnwavyrpgvgrygiedipvekcthiiysfigvtegnsevliidpeldvdkng 82
          |::||||:||| || | :|:   |:    |||:||:| |:|: : : |    | : |
Qy      1 AKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTN-H-Q-LST-TEWN-DETL 55

Db     83 frnftslrsshpsvkfmvavggwaegsskyshmvaqkstrmsfirsvvsflkkydfdgld 142
          :::| :|:  :| :| ::|:|||  |: |:: |||  : | :|: | : ||:|| ||||
Qy     56 YQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLD 115

Db    143 ldweypgaadrggsfsdkdkflylvqelrraf---irv-gkgwel-taavplanfrlmeg 197
          |||||||:  :|::  ||::|  |||:|  ||    :  || : | :|||| :   :  |
Qy    116 LDWEYPGS--QGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAG 173

Db    198 yhvpelcqeldaihvmsydlrgnwagfadvhsplykrphdqwayeklnvndglhlweekg 257
          | |  : |:|| :::|:|||::|:|   :  :|||||  : |   |||: ::: | :||
Qy    174 YEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKG 233

Db    258 cpsnklvvgipfygrsftlsagnnnyglgtfinkeagggdpapytnatgfwayyeictev 317
          |::||::|:| ||||||:::::   :|:     :|:| |:|:|:  |  ||||:|:
Qy    234 TPASKLILGMPTYGRSFTLASSSDT-RVGA--P-ATGSGTPGPFTKEGGMLAYYEVCS-- 287

Db    318 dkddsgwtkkwdeqgkcpyaykgtqwvgyedprsveikmnwikqkgylgamtwaidmddf 377
          |  :  || : :: | || ::  |||::|  |   |:: :|||  ||| ||:|:|||
Qy    288 WKG-A--TKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDF 344

Db    378 qgl-cge-knplikilhkhms-sy 398
          |: | : : |||  |:  :| :|
Qy    345 AGFSCNQGRYPLIQTLRQELSLPY 368


RESULT   11
ENTRY            A38368     #type complete
TITLE            chitinase (EC 3.2.1.14) precursor - Bacillus circulans
ORGANISM         #formal_name Bacillus circulans
DATE             28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change
                   03-May-1996
ACCESSIONS       A38368
REFERENCE        A38368
   #authors      Watanabe, T.; Suzuki, K.; Oyanagi, W.; Ohnishi, K.; Tanaka,
                   H.
   #journal      J. Biol. Chem. (1990) 265:15659-15665
   #title        Gene cloning of chitinase A1 from Bacillus circulans WL-12
                   revealed its evolutionary relationship to Serratia
                   chitinase and to the type III homology units of
                   fibronectin.
   #cross-references MUID:90368776
   #accession    A38368
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-699 ##label WAT
      ##cross-references GB:J05599
CLASSIFICATION   #superfamily fibronectin type III repeat homology
KEYWORDS         glycosidase; hydrolase
SUMMARY          #length 699  #molecular-weight 73677  #checksum 6757

  Query Match           25.7%;  Score 698;  DB 10;  Length 699;
  Best Local Similarity  40.1%;  Pred. No. 2.65e-101;
  Matches   119;  Conservative   73;  Mismatches 85;  Indels  20;  Gaps  18;

Db    144 qlnklkqtnpnlktiisvggwtw-snrfsdvaataatrevfansavdflrkynfdgvdld 202
          ::| ||  ||:|||::::|||   : :|:|: |||  |: | |||: |||||:|||:|||
Qy     58 EFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 117

Db    203 weypvsggldgnskrpedkqnytlllskirekldaagavdgk-kylltiasgasatya-a 260
          ||||  |: :| |  : ||: :| |:  : : ::  :   || : ||: |  |: ||  |
Qy    118 WEYP--GS-QG-S-PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDA 172

Db    261 ntelakiaaivdwinimtydfngawqkisahnaplnydpaasaagvpdantfnvaagaqg 320
            |: |||  :| :|:|:|||:|:|:|:::||:|| |      :| : | ::|| |: |
Qy    173 GYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPL-YKRQE-ESGAA-A-SLNVDAAVQQ 228

Db    321 hldagvpaaklvlgvpfygrgwd-gcaqagn-gqyqtctggssvgtweagsfdfydlean 378
          |: | ||:||:||:| |||:   : :     |  | :| ::  | |:| : :|::  :
Qy    229 WLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEV-CS 287

Db    379 yinkngytrywndtakvpylynasnkrfisyddaesvgyktayikskglggamfwel 435
            | |: : :  ||||::   | : :::|| ||   | :|:| |||||||  | |
Qy    288 W--K-GATKQRIQDQKVPYIFRD-NQW-VGFDDVESFKTKVSYLKQKGLGGAMVWAL 339


RESULT   12
ENTRY            S51591     #type complete
TITLE            chitinase (EC 3.2.1.14)/lysozyme (EC 3.2.1.17) PZ precursor,
                   pathogenesis-related - common tobacco
ORGANISM         #formal_name Nicotiana tabacum #common_name common tobacco
DATE             15-Jul-1995 #sequence_revision 08-Sep-1995 #text_change
                   03-May-1996
ACCESSIONS       S51591; S51632; S43119
REFERENCE        S51591
   #authors      Heitz, T.; Segond, S.; Kauffmann, S.; Geoffroy, P.; Prasad,
                   V.; Brunner, F.; Fritig, B.; Legrand, M.
   #journal      Mol. Gen. Genet. (1994) 245:246-254
   #title        Molecular characterization of a novel tobacco
                   pathogenesis-related (PR) protein: a new plant
                   chitinase/lysozyme.
   #accession    S51591
      ##molecule_type mRNA
      ##residues      1-378 ##label HEI
      ##cross-references EMBL:X78325
```

```
     ##experimental_source cv. Samsun NN
   #accession    S51632
     ##molecule_type protein
     ##residues      31-36;87-112;252-275;282-305;337-371 ##label HEW
CLASSIFICATION   #superfamily Streptomyces chitinase chi40
KEYWORDS         glycosidase; hydrolase
FEATURE
   1-25               #domain signal sequence #status predicted #label SIG\
   26-371             #product chitinase/lysozyme PZ #status predicted #label
                        MAT\
   371-378            #domain carboxyl-terminal propeptide #status
                        experimental #label PRO
SUMMARY          #length 378  #molecular-weight 42019  #checksum 6689

  Query Match           21.9%;  Score 594;  DB 5;  Length 378;
  Best Local Similarity  32.7%;  Pred. No. 3.36e-82;
  Matches   106;  Conservative   87;  Mismatches 108;  Indels  23;  Gaps  18;

Db     43 nnidstlfthlfcafadlnpqsnqliispenqds-fsqftstvqrknpsvktflsiaggr 101
          :::|::| |||: ||| :   ::||  :  |::: : :| : : ::|| :||:|:| ||:
Qy     23 KDLDPSLCTHLIYAFAGMT--NHQLSTTEWNDETLYQEFNG-LKKMNPKLKTLLAI-GGW 78

Db    102 a-dttaygimarqpnsrksfidssirlarqfgfhgldldweyp--l-s-atdmtnlgill 156
            |  :  |   :|:| :|::|:||: | ::| |||||||||    | | |:  :  |:
Qy     79 NFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYPGSQGSPAVDKERFTTLV 138

Db    157 newrtainmearnsgraallltaavsysprvnglnypvesvarnlnwinlmaydfygpnw 216
           :   |:  ||: ||:  |||:|||: :       | |: :|:||: :|||||| |: |
Qy    139 QDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGS-W 197

Db    217 spsqtnshaqlfdpvnh--isgs---dg-inawiqagvptkklvlgipfygyawrlvnpn 270
             |  :: |:   :    ::|   |: :  |:| | |: ||:||:| || :  | :::
Qy    198 EKV-TGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSS 256

Db    271 ihdlrapaagksnvga-vddgsm-tynrirdyivqsrattvynativgdycysgsnwisy 328
           : |||:| :  |:   :|:| :|  :      ::   : :   |  | :  : |:::
Qy    257 DTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQR-IQDQK-VP-YIFRDNQWVGF 313

Db    329 ddtqsvrnkvnyvkgrgllgyfaw 352
          || :| : ||:|:| :|| |   |
Qy    314 DDVESFKTKVSYLKQKGLGGAMVW 337


RESULT   13
ENTRY            S47133     #type complete
TITLE            chitinase (EC 3.2.1.14) - fungus (Trichoderma harzianum)
ORGANISM         #formal_name Trichoderma harzianum
DATE             23-Nov-1994 #sequence_revision 03-Aug-1995 #text_change
                   10-May-1996
ACCESSIONS       S47133
REFERENCE        S47133
   #authors      Carsolio, C.; Gutierrez, A.; Jimenez, B.; van Montagu, M.;
                   Herrera-Estrella, A.
   #submission   submitted to the EMBL Data Library, June 1994
   #description  Characterization of ech-42, a Trichoderma harzianum
                   endochitinase gene expressed during mycoparasitism.
   #accession    S47133
     ##status          preliminary
     ##molecule_type DNA
     ##residues      1-424 ##label CAR
     ##cross-references EMBL:X79381
GENETICS
   #introns      48/2; 81/2; 98/1
CLASSIFICATION   #superfamily Streptomyces chitinase chi40
KEYWORDS         glycosidase; hydrolase
SUMMARY          #length 424  #molecular-weight 46298  #checksum 8773

  Query Match           21.4%;  Score 582;  DB 5;  Length 424;
  Best Local Similarity  36.0%;  Pred. No. 5.18e-80;
  Matches   102;  Conservative   69;  Mismatches  95;  Indels  17;  Gaps  13;

Db    112 qlfk-lkkanrnlkvmlsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidv 169
          | |: ||| | :|| :|:|||| :| :|    :|  ||  | ::|| |::  :|||:|:

Qy     57 QEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 116

Db    170 dweyp-addtqatnmvlllkeirsqldayaaqy-apg-yhfllsiaapagpehy-sflhm 225
          ||||| :: : | ::  :   ::  :|:  :  ::|  ::||| | |||     :   :
Qy    117 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 176

Db    226 sdlgqvldyvnlmaydyagswssysghdanlfanpsnpnssp-yntdqaikdyikggvpa 284
            ::| ||:|||||||: |||   :||:: |:    :: :::  | | |: : :  | ||
Qy    177 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 236

Db    285 skivlgmpiygraf--estgg--igqtysgigs-gsw--engiwdy-kvlpkagatvgyd 336
          ||::|||| |||:|   |::   :|   :| |: |:   | |:  |  | :  ||| |
Qy    237 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQ-- 294

Db    337 svaqayysydpsskelisfdtpdmintkvsylknlglggsmfw 379
          : :   :|    :: ::||  : ::|||||||  ||||:| |
Qy    295 RIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT   14
ENTRY            JQ1975     #type complete
TITLE            chitinase (EC 3.2.1.14) 1 precursor - imperfect fungus
                   (Aphanocladium album)
ORGANISM         #formal_name Aphanocladium album
DATE             03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change
                   03-May-1996
ACCESSIONS       JQ1975
REFERENCE        JQ1975
   #authors      Blaiseau, P.L.; Lafay, J.F.
   #journal      Gene (1992) 120:243-248
   #title        Primary structure of a chitinase-encoding gene (chi1) from
                   the filamentous fungus Aphanocladium album: Similarity to
                   bacterial chitinases.
   #accession    JQ1975
     ##molecule_type mRNA
     ##residues      1-423 ##label BLA
     ##cross-references GB:X64104
COMMENT     This enzyme is essential for the degradation of insoluble chitin.
GENETICS
   #gene         chi1
   #introns      47/2; 80/2; 97/1
CLASSIFICATION   #superfamily Streptomyces chitinase chi40
KEYWORDS         glycosidase; hydrolase
FEATURE
   1-34               #domain signal sequence #status predicted #label SIG\
   35-423             #product chitinase 1 #status predicted #label MAT
SUMMARY          #length 423  #molecular-weight 46072  #checksum 8802

  Query Match           21.2%;  Score 576;  DB 5;  Length 423;
  Best Local Similarity  36.2%;  Pred. No. 6.42e-79;
  Matches   101;  Conservative   65;  Mismatches  95;  Indels  18;  Gaps  16;

Db    115 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 173
          ||| | ::| :|:|||| :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    174 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 228
           :::: | ::  :   |:   ::: :  : | ::||| | ||| : |    :  : ::
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 180

Db    229 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 287
            ||::|||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    288 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 339
          |||| |||||   :  :  :| | | |: |:   | |:  |  :    |||    :  |
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 300

Db    340 gcysydpstkelisfdtpamistkvswlkgkglggsmfw 378
            | :    :: ::||    : |||| || |||||:| |
Qy    301 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337
```

```
RESULT   15
ENTRY            S51369     #type complete
TITLE            chitinase - fungus (Trichoderma harzianum)
ORGANISM         #formal_name Trichoderma harzianum
DATE             15-Jul-1995 #sequence_revision 19-Apr-1996 #text_change
                   03-May-1996
ACCESSIONS       S51369
REFERENCE        S51369
   #authors      Garcia, I.; Lora, J.M.; de la Cruz, J.; Benitez, T.; Llobell,
                   A.; Pintor-Toro, J.A.
   #journal      Curr. Genet. (1994) 27:83-89
   #title        Cloning and characterization of a chitinase (CHIT42) cDNA
                   from the mycoparasitic fungus Trichoderma harzianum.
   #accession    S51369
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-423 ##label GAR
SUMMARY          #length 423  #molecular-weight 46056  #checksum 2000

  Query Match           20.6%;  Score 561;  DB 12;  Length 423;
  Best Local Similarity  35.7%;  Pred. No. 3.43e-76;
  Matches   101;  Conservative   71;  Mismatches  94;  Indels  17;  Gaps  13;

Db    111 qlfk-vkkanrglkvllsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidi 168
          | |: :|| |  || ||:|||| :| :|    :|  ||  | ::|| |::  :|||:|:
Qy     57 QEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 116

Db    169 dweyp-adatqasnmilllkevrsqrdayaaqy-apg-yhflltiaapagkdny-sklrl 224
          ||||| :::: | ::  :   |:   :|:  :  ::|  ::||: | |||     :   :
Qy    117 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 176

Db    225 adlgqvldyinlmaydyagsfspltghdanlfnnpsnpnatp-fntdsavkdyinggvpa 283
            ::| ||::|||||: ||   :|||:: |::   :: |:: :| |:|| : :  | ||
Qy    177 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 236

Db    284 nkivlgmpiygrsf--qntag--igqtyngvgs-gsw--eagiwdykalpk-agatvqyd 335
          :|::|||| |||||   :::   :|    | |: |:   |:|:  |  :    ||| |
Qy    237 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRI 296

Db    336 svakgyysynsatkelisfdtpdmintkvaylkslglggsmfw 378
            | | :    :: ::||  : ::|||:|||  ||||:| |
Qy    297 QDQKVPYIFRD--NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


Search completed: Fri Jun 27 17:06:56 1997
Job time : 134 secs.
```

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 27 17:07:13 1997; MasPar time 12.20 Seconds
648.449 Million cell updates/sec

Tabular output not generated.

Title: >US-08-663-618A-15
Description: (1-373) from US08663618A.pep
Perfect Score: 2717
Sequence: 1 AKLVCYFTNWAQYRQGEARF.....YPLIQTLRQELSLPYLSSGT 373

Scoring table: PAM 150
Gap 11

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot34
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11

Statistics: Mean 48.971; Variance 85.460; scale 0.573

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1488	54.8	383	4	GP39_HUMAN	CARTILAGE GLYCOPROTEI	0.00e+00
2	1086	40.0	504	2	CHIT_BRUMA	ENDOCHITINASE PRECURS	1.03e-218
3	1015	37.4	554	2	CHIT_MANSE	ENDOCHITINASE PRECURS	7.72e-202
4	698	25.7	699	2	CHI1_BACCI	CHITINASE A1 PRECURSO	2.43e-127
5	576	21.2	423	2	CHI1_APHAL	CHITINASE 1 PRECURSOR	3.13e-99
6	561	20.6	423	2	CHI4_TRIHA	42 KD ENDOCHITINASE P	8.30e-96
7	546	20.1	427	2	CHI1_COCIM	ENDOCHITINASE 1 PRECU	2.16e-92
8	437	16.1	563	2	CHIA_SERMA	CHITINASE A PRECURSOR	7.14e-68
9	395	14.5	551	2	CHIT_NPVAC	PROBABLE ENDOCHITINAS	1.33e-58
10	345	12.7	820	2	CHIA_ALTSO	CHITINASE A PRECURSOR	9.66e-48
11	337	12.4	499	2	CHIB_SERMA	CHITINASE B PRECURSOR	5.03e-46
12	301	11.1	619	2	CHIT_STRLI	CHITINASE C PRECURSOR	2.19e-38
13	287	10.6	610	2	CHIT_STRPL	CHITINASE 63 PRECURSO	1.86e-35
14	263	9.7	1146	5	KTXA_KLULA	KILLER TOXIN ALPHA AN	1.67e-30
15	172	6.3	36	4	GP39_BOVIN	CARTILAGE GLYCOPROTEI	8.53e-13
16	172	6.3	83	3	EDOS_PAPAN	ESTRADIOL-DEPENDENT O	8.53e-13
17	121	4.5	597	2	CHIX_STROI	EXOCHITINASE 1 PRECUR	3.64e-04
18	108	4.0	452	4	HEMF_RHOSH	COPROPORPHYRINOGEN II	3.11e-02
19	110	4.0	497	2	CHID_BACCI	CHITINASE D PRECURSOR	1.60e-02
20	104	3.8	422	10	UMUC_ECOLI	UMUC PROTEIN.	1.14e-01
21	100	3.7	364	11	YBR5_YEAST	HYPOTHETICAL 40.9 KD	4.02e-01
22	101	3.7	1045	7	PRTT_SERMA	EXTRACELLULAR SERINE	2.94e-01
23	101	3.7	1070	1	AGLU_CANTS	ALPHA-GLUCOSIDASE PRE	2.94e-01
24	99	3.6	398	9	SPEB_STRPY	EXOTOXIN TYPE B PRECU	5.48e-01
25	98	3.6	492	9	SYK_THETH	LYSYL-TRNA SYNTHETASE	7.44e-01
26	97	3.6	563	5	K2CB_HUMAN	KERATIN, TYPE II CYTO	1.01e+00
27	97	3.6	563	5	K2CA_HUMAN	KERATIN, TYPE II CYTO	1.01e+00
28	97	3.6	563	5	K2CC_HUMAN	KERATIN, TYPE II CYTO	1.01e+00
29	99	3.6	958	1	AMYG_SCHOC	GLUCOAMYLASE 1 PRECUR	5.48e-01
30	98	3.6	1045	7	PRTS_SERMA	EXTRACELLULAR SERINE	7.44e-01
31	96	3.5	297	5	HMX1_HUMAN	HOMEOBOX PROTEIN MSX-	1.36e+00
32	94	3.5	344	4	GCH2_PICGU	GTP CYCLOHYDROLASE II	2.47e+00
33	95	3.5	353	11	YQHT_BACSU	PUTATIVE PEPTIDASE IN	1.84e+00
34	94	3.5	404	6	MOEA_HAEIN	MOLYBDOPTERIN BIOSYNT	2.47e+00
35	95	3.5	492	7	PDI_SCHPO	PUTATIVE PROTEIN DISU	1.84e+00
36	96	3.5	554	5	HYES_RAT	SOLUBLE EPOXIDE HYDRO	1.36e+00
37	95	3.5	554	5	HYES_MOUSE	SOLUBLE EPOXIDE HYDRO	1.84e+00
38	95	3.5	563	5	K2CE_HUMAN	KERATIN, TYPE II CYTO	1.84e+00
39	95	3.5	563	5	K2CF_HUMAN	KERATIN, TYPE II CYTO	1.84e+00
40	94	3.5	993	11	YAJ1_SCHPO	PUTATIVE FAMILY 31 GL	2.47e+00
41	95	3.5	1333	10	VGR1_MOUSE	VASCULAR ENDOTHELIAL	1.84e+00
42	93	3.4	267	3	EBAG_FLASP	ENDO-BETA-N-ACETYLGLU	3.31e+00
43	93	3.4	295	11	YHO8_YEAST	HYPOTHETICAL 35.1 KD	3.31e+00
44	92	3.4	510	1	BMP6_MOUSE	BONE MORPHOGENETIC PR	4.43e+00
45	92	3.4	557	3	ERK3_HUMAN	EXTRACELLULAR SIGNAL-	4.43e+00

ALIGNMENTS

RESULT 1
ID GP39_HUMAN STANDARD; PRT; 383 AA.
AC P36222; P30923;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE CARTILAGE GLYCOPROTEIN-39 PRECURSOR (GP-39) (39 KD SYNOVIAL PROTEIN).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=CARTILAGE;
RX MEDLINE; 94064658.
RA HAKALA B.E., WHITE C., RECKLIES A.D.;
RL J. BIOL. CHEM. 268:25803-25810(1993).
RN [2]
RP SEQUENCE OF 22-45.
RX MEDLINE; 90328983.
RA NYIRKOS P., GOLDS E.E.;
RL BIOCHEM. J. 269:265-268(1990).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE CAPACITY OF CELLS TO
CC RESPOND TO AND COPE WITH CHANGES IN THEIR ENVIRONMENT.
CC -!- TISSUE SPECIFICITY: PRESENT IN ARTICULAR CHONDROCYTES, SYNOVIAL
CC CELLS AS WELL AS IN LIVER. UNDETECTABLE IN MUSCLE TISSUES, LUNG,
CC PANCREAS, MONONUCLEAR CELLS, OR FIBROBLASTS.
CC -!- PTM: GLYCOSYLATED.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
DR EMBL; M80927; G348912; -.
DR PIR; S10677; S10677.
DR PIR; A33162; A33162.
DR PROSITE; PS01095; CHITINASE_18.
KW GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 21
FT CHAIN 22 383 CARTILAGE GLYCOPROTEIN-39.
FT CARBOHYD 60 60 POTENTIAL.
SQ SEQUENCE 383 AA; 42613 MW; D0A5280F CRC32;

Query Match 54.8%; Score 1488; DB 4; Length 383;
Best Local Similarity 53.3%; Pred. No. 0.00e+00;
Matches 195; Conservative 83; Mismatches 79; Indels 9; Gaps 5;

```
Db     23 klvcyytswsqyregdgscfpdaldrflcthiiysfanisndhidtwewndvtlygmlnt 82
          |||||:|:|:|||:|::  :|  ||  ||||:||:|| ::| :: | |||| |||  :|
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTTEWNDETLYQEFNG 61
```

```
Db     83 lknrnpnlktllsvggwnfgsqrfskiasntqsrrtfiksvppflrthgfdgldlawlyp 142
          ||: ||:|||||::|||||:|:|: : : : :|:||::|   |||  :|||||| | ||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    143 grr-----dkqhfttlikemkaefikeaq-pgkkqlllsaalsagkvtidssydiakisq 196
          | :     ||::||||: ::   |  ||| :|| :|||||::|| :|::|:: ||:|
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQ 181

Db    197 hldfisimtydfhgawrgttghhsplfrgqedaspdrfsntdyavgymlrlgapasklvm 256
          :|||:::|:|||||:|   |||:|||:: ||::::    | | ||   |: |:|||||::
Qy    182 NLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLIL 241

Db    257 giptfgrsftlass-etgvgapisgpgipgrftkeagtlayyeicdflrgatvhrtlgqq 315
          |:||:|||||||||| :| |||| :|:| || ||||:| ||||:|   :||| :|   |
Qy    242 GMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRIQDQK 300

Db    316 vpyatkgnqwvgyddqesvkskvqylkdrqlagamvwaldlddfqgsfcgqdlrfpltna 375
          |||  : |||||:|| || |:|| |||:: |:|||||||||||| |  | |  |:||  :
Qy    301 VPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQG-RYPLIQT 359

Db    376 ikdala 381
          ::: |:
Qy    360 LRQELS 365


RESULT    2
ID   CHIT_BRUMA     STANDARD;      PRT;   504 AA.
AC   P29030;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   ENDOCHITINASE PRECURSOR (EC 3.2.1.14) (MF1 ANTIGEN).
OS   BRUGIA MALAYI.
OC   EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; SPIRURIDA.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE; 92179220.
RA   FUHRMAN J.A., LANE W.S., SMITH R.F., PIESSENS W.F., PERLER F.B.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 89:1548-1552(1992).
CC   -!- FUNCTION: THE MF1 ANTIGEN IS A MICROFILARIAL CHITINASE, WHICH MAY
CC       FUNCTION TO DEGRADE CHITIN-CONTAINING STRUCTURES IN THE MICRO-
CC       FILARIA OR IN ITS MOSQUITO VECTOR DURING PARASITE DEVELOPMENT AND
CC       TRANSMISSION.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- PTM: O-GLYCOSYLATED.
CC   -!- KNOWN TO BIND CALCIUM.
CC   -!- DEVELOPMENTAL STAGE: THE APPEARANCE OF THE MF1 ANTIGEN CORRESPOND
CC       WITH THE ONSET OF THE PARASITE'S ABILITY TO INFECT THE MOSQUITO.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; M73689; G156064; -.
DR   PIR; A38221; A38221.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ANTIGEN;
KW   REPEAT; GLYCOPROTEIN; CALCIUM-BINDING.
FT   SIGNAL        1     22
FT   CHAIN        23    504       ENDOCHITINASE.
FT   DOMAIN       23    400       CATALYTIC.
FT   DOMAIN      401    450       SER/THR-RICH (LINKER).
FT   DOMAIN      407    448       3 X 14 AA APPROXIMATE TANDEM REPEATS.
FT   ACT_SITE    148    148       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   504 AA;  55971 MW;  4DA7E5EB CRC32;

  Query Match           40.0%;  Score 1086;  DB 2;  Length 504;
  Best Local Similarity  42.8%;  Pred. No. 1.03e-218;
  Matches   158;  Conservative   94;  Mismatches 102;  Indels  15;  Gaps  11;

Db     27 cyytnwaqyrdgegkflpgnipnglcthilyafakvdelgdskpfewndedtewskgmys 86
          ||:|||||||:||::||| ::  :||||::|||| : :   |   ||||| |     :|
Qy      5 CYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTNHQLSTT-EWNDE-T-----LYQ 57

Db     87 avtklretnpglkvllsyggynfgsaiftgiaksaqkterfiksaiaflrknnfdgfdld 146
             |:  || || ||: || |||:  || :  :| : : |::||| |||| :|||:|||
Qy     58 EFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 117

Db    147 weyp-v-gv-a-e-eh-aklveamktafveeaktsgkqrllltaavsagkgtidgsynve 200
          ||||   |  | : |: : ||: :  || :|| ||||:||||:|||:||   :|::|:|:
Qy    118 WEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEVD 177

Db    201 slgknfdllflmsydlhgsweknvdlhgklhptkgevsgigifntefaadywaskgmpke 260
          :: |:|:: ||:||:|||||    :: |     | :: : :| : | : |  || |
Qy    178 KIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPAS 237

Db    261 kiiigipmyaqgwtldnpsetaigaaasrpssasktnpaggtasyweickylkeggketv 320
          |:|:|:| |::: || ::|:| :||:|: :::::     ||  :| |:|   | : |: :
Qy    238 KLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSW-KGATKQRI 296

Db    321 hqegvgaymvkgdqwygydneetirikmkwlkekgyggafiwaldfddftgkscgkgpyp 380
          ::: | :|: : :|| |:|: |::: |:  ||:|| ||| :||||:|||:| ||  | ||
Qy    297 QDQKV-PYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAGFSCNQGRYP 355

Db    381 llnaissel 389
          |: ::´ ||
Qy    356 LIQTLRQEL 364


RESULT    3
ID   CHIT_MANSE     STANDARD;      PRT;   554 AA.
AC   P36362;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
OS   MANDUCA SEXTA (TOBACCO HAWKMOTH) (TOBACCO HORNWORM).
OC   EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; LEPIDOPTERA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93357793.
RA   KRAMER K.J., CORPUZ L., CHOI H.K., MUTHUKRISHNAN S.;
RL   INSECT BIOCHEM. MOL. BIOL. 23:691-701(1993).
CC   -!- FUNCTION: DIGEST CHITIN IN THE EXOSKELETON DURING THE MOLTING
CC       PROCESS.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- DEVELOPMENTAL STAGE: HIGH LEVELS SEEN IN THE EPIDERMIS ON DAY 0,
CC       BUT RAPIDLY DISAPPEARS AND IS UNDETECTED ON DAYS 1-4 OF FIFTH
CC       INSTAR. IT REAPPEARS ON DAY 5 AND PEAKS ON DAY 7 AFTER WHICH A
CC       RAPID DECLINE IS SEEN. IN THE GUT IS DETECTED ON DAY 6 WITH LOWER
CC       LEVELS SEEN ON DAYS 0, 7 AND 8.
CC   -!- TISSUE SPECIFICITY: EPIDERMIS AND GUT.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; U02270; G406049; -.
DR   EMBL; S64757; -; NOT_ANNOTATED_CDS.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    554       ENDOCHITINASE.
FT   DOMAIN      396    453       SER/THR-RICH.
FT   ACT_SITE    146    146       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD     85     85       POTENTIAL.
FT   CARBOHYD    303    303       POTENTIAL.
FT   CARBOHYD    407    407       POTENTIAL.
FT   CARBOHYD    545    545       POTENTIAL.
SQ   SEQUENCE   554 AA;  62203 MW;  FA87F8AD CRC32;

  Query Match           37.4%;  Score 1015;  DB 2;  Length 554;
  Best Local Similarity  40.4%;  Pred. No. 7.72e-202;
  Matches   155;  Conservative   98;  Mismatches 107;  Indels  24;  Gaps  18;

Db     23 arivcyfsnwavyrpgvgrygiedipvekcthiiysfigvtegnsevliidpeldvdkng 82
          |::||||:||| || | :|:   |:    |||:||:| |:|: : : |    | : |
```

```
Qy       1 AKLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTN-H-Q-LST-TEWN-DETL 55

Db      83 frnftslrsshpsvkfmvavggwaegsskyshmvaqkstrmsfirsvvsflkkydfdgld 142
           :::| :|:  :| :| ::|:|||  |: |:: |||  : | :|: | : ||:|| |||||
Qy      56 YQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLD 115

Db     143 ldweypgaadrggsfsdkdkflylvqelrraf---irv-gkgwel-taavplanfrlmeg 197
           |||||||:  :|::  ||::|  |||:|  ||    :  || : | :|||| :   :  |
Qy     116 LDWEYPGS--QGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAG 173

Db     198 yhvpelcqeldaihvmsydlrgnwagfadvhsplykrphdqwayeklnvndglhlweekg 257
           | |  : |:|| :::|:|||::|:|   :  :|||||  :  |   |||: ::: | :||
Qy     174 YEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKG 233

Db     258 cpsnklvvgipfygrsftlsagnnnyglgtfinkeagggdpapytnatgfwayyeictev 317
           |::||::|:| |||||||:::::   :|:    :|:| |:|:|:  |  ||||:|:
Qy     234 TPASKLILGMPTYGRSFTLASSSDT-RVGA--P-ATGSGTPGPFTKEGGMLAYYEVCS-- 287

Db     318 dkddsgwtkkwdeqgkcpyaykgtqwvgyedprsveikmnwikqkgylgamtwaidmddf 377
           |  :  || : :: | || ::  ||||::|  |   |:: :|||| |||  ||:|:|||
Qy     288 WKG-A--TKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDF 344

Db     378 qgl-cge-knplikilhkhms-sy 398
           |: | : : |||  |:  :| :|
Qy     345 AGFSCNQGRYPLIQTLRQELSLPY 368


RESULT    4
ID   CHI1_BACCI     STANDARD;      PRT;   699 AA.
AC   P20533;
DT   01-FEB-1991 (REL. 17, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE A1 PRECURSOR (EC 3.2.1.14).
GN   CHIA1.
OS   BACILLUS CIRCULANS.
OC   PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WL-12;
RX   MEDLINE; 90368776.
RA   WATANABE T., SUZUKI K., OYANAGI W., OHNISHI K., TANAKA H.;
RL   J. BIOL. CHEM. 265:15659-15665(1990).
RN   [2]
RP   MUTAGENESIS.
RC   STRAIN=WL-12;
RX   MEDLINE; 93366760.
RA   WATANABE T., KOHORI K., MIYASHITA K., FUJII T., SAKAI H.,
RA   UCHIDA M., TANAKA H.;
RL   J. BIOL. CHEM. 268:18567-18572(1993).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; M57601; G142688; -.
DR   PIR; A38368; A38368.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; REPEAT.
FT   SIGNAL        1     41
FT   CHAIN        42    699       CHITINASE A1.
FT   DOMAIN       42    460       CATALYTIC.
FT   DOMAIN      465    549       FIBRONECTIN TYPE-III (R-1).
FT   DOMAIN      560    644       FIBRONECTIN TYPE-III (R-2).
FT   ACT_SITE    204    204       PROTON DONOR (PROBABLE).
FT   MUTAGEN     200    200       D->N: DECREASE IN ACTIVITY.
FT   MUTAGEN     200    200       D->E: NO CHANGE IN ACTIVITY.
FT   MUTAGEN     204    204       E->D,Q: LOSS OF ACTIVITY.
SQ   SEQUENCE   699 AA;  73677 MW;  5A1777CC CRC32;

  Query Match           25.7%;  Score 698;  DB 2;  Length 699;
  Best Local Similarity 40.1%;  Pred. No. 2.43e-127;
  Matches  119;  Conservative   73;  Mismatches  85;  Indels  20;  Gaps  18;

Db     144 qlnklkqtnpnlktiisvggwtw-snrfsdvaataatrevfansavdflrkynfdgvdld 202
           ::| ||  ||:|||::::|||   : :|:|: |||  |: | |||: |||||:|||:|||
Qy      58 EFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLD 117

Db     203 weypvsggldgnskrpedkqnytlllskirekldaagavdgk-kylltiasgasatya-a 260
           ||||  |: :| |  : ||: :| |:  : : ::  :   || : ||: |  |: ||  |
Qy     118 WEYP--GS-QG-S-PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDA 172

Db     261 ntelakiaaivdwinimtydfngawqkisahnaplnydpaasaagvpdantfnvaagaqg 320
             |: |||  :| :|:|:|||:|:|:|:::||:|| |       :| : | ::|| |: |
Qy     173 GYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPL-YKRQE-ESGAA-A-SLNVDAAVQQ 228

Db     321 hldagvpaaklvlgvpfygrgwd-gcaqagn-gqyqtctggssvgtweagsfdfydlean 378
           |: | ||:||:||:| |||:   : :     |   | :| ::  | |:| : :|::  :
Qy     229 WLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEV-CS 287

Db     379 yinkngytrywndtakvpylynasnkrfisyddaesvgyktayikskglggamfwel 435
           | | |: : :  ||||::   | : :::|| ||   | :|:| |||||| | |
Qy     288 W--K-GATKQRIQDQKVPYIFRD-NQW-VGFDDVESFKTKVSYLKQKGLGGAMVWAL 339


RESULT    5
ID   CHI1_APHAL     STANDARD;      PRT;   423 AA.
AC   P32470;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   CHITINASE 1 PRECURSOR (EC 3.2.1.14).
GN   CHI1.
OS   APHANOCLADIUM ALBUM.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ETHM 483;
RX   MEDLINE; 93013040.
RA   BLAISEAU P.-L., LAFAY J.-F.;
RL   GENE 120:243-248(1992).
RN   [2]
RP   SEQUENCE OF 35-57.
RX   MEDLINE; 92136437.
RA   BLAISEAU P.-L., KUNZ C., GRISON R., BERTHEAU Y., BRYGOO Y.;
RL   CURR. GENET. 21:61-66(1992).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; X64104; G429026; -.
DR   PIR; JQ1975; JQ1975.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ZYMOGEN.
FT   SIGNAL        1     22       POTENTIAL.
FT   PROPEP       23     34       POTENTIAL.
FT   CHAIN        35    423       CHITINASE 1.
FT   ACT_SITE    171    171       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   423 AA;  46058 MW;  88456E6E CRC32;

  Query Match           21.2%;  Score 576;  DB 2;  Length 423;
  Best Local Similarity 36.2%;  Pred. No. 3.13e-99;
  Matches  101;  Conservative   65;  Mismatches  95;  Indels  18;  Gaps  16;

Db     115 lkkqnrnmkvmlsiggwtwst-nfpaaassaatrktfaqsavgfmkdwgfdgididweyp 173
           ||| | ::| :|:|||| :| :|    ::|  | ||  ||: |::  :|||:|:|||||
Qy      62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db     174 -adatqaqnmvlllqavrseldsyaaqy-akg-hhfllsiaapagpdnynklkfa-e-lg 228
            :::: | ::  :  |:   ::: :  : |  ::||| | ||| : |    :  : ::
Qy     122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 180

Db     229 kvldyinlmaydyagswsnytghdaniyanpqnpnatp-yntddavqayinggvpankiv 287
```

```
          ||::|||||||: ||| : |||:: :|   ::: |::  | | |||  :  | ||:|::
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    288 lgmpiygrsfq-qt--eg-igkpyngigs-gsw--engiwdykalpk-agatvkcddtak 339
          |||| |||||   :  :| |  | |: |:   | |:  |  :    |||    :  |
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 300

Db    340 gcysydpstkelisfdtpamistkvswlkgkglggtmfw 378
            | :    :: ::||    : |||| || |||||:| |
Qy    301 VPYIF-RD-NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT    6
ID   CHI4_TRIHA     STANDARD;      PRT;   423 AA.
AC   P48827;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   42 KD ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
GN   CHIT42.
OS   TRICHODERMA HARZIANUM.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 35-52; 93-107; 371-385 & 397-414.
RX   MEDLINE; 95269313.
RA   GARCIA I., LORA J.M., LA CRUZ J., BENITEZ T., LLOBELL A.,
RA   PINTOR-TORO J.A.;
RL   CURR. GENET. 27:83-89(1994).
CC   -!- FUNCTION: MORPHOGENETIC ROLE DURING APICAL GROWTH,CELL DIVISION
CC       AND DIFFERENTIATION (CELL WALL MORPHOGENESIS). ANTIFUNGAL
CC       AGENT.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- INDUCTION: SPECIFICALLY INDUCED BY CHITIN AND IS CATABOLITE
CC       REPRESSED.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; S78423; G999376; -.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ZYMOGEN;
KW   CHITIN-BINDING.
FT   SIGNAL        1     22       POTENTIAL.
FT   PROPEP       23     34
FT   CHAIN        35    423       42 KD ENDOCHITINASE.
FT   ACT_SITE    171    171       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    218    218       POTENTIAL.
SQ   SEQUENCE   423 AA;  46056 MW;  899DA50A CRC32;

  Query Match           20.6%;  Score 561;  DB 2;  Length 423;
  Best Local Similarity  35.7%;  Pred. No. 8.30e-96;
  Matches   101;  Conservative   71;  Mismatches  94;  Indels  17;  Gaps  13;

Db    111 qlfk-vkkanrglkvllsiggwtwst-nfpsaastdanrknfaktaitfmkdwgfdgidi 168
          | |: :|| |  || ||:|||| :| :|    :|  ||  | ::|| |::  :|||:|:
Qy     57 QEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDL 116

Db    169 dweyp-adatqasnmilllkevrsqrdayaaqy-apg-yhfltiaapagkdny-sklrl 224
          ||||| :::: | ::  :   |:   :|:  :  ::|  ::||: | |||     :   :
Qy    117 DWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVDAGYEV 176

Db    225 adlgqvldyinlmaydyagsfspltghdanlfnnpsnpnatp-fntdsavkdyinggvpa 283
            ::| ||::|||||: ||   :|||:: |::   :: |:: :| |:|| : :  | ||
Qy    177 DKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPA 236

Db    284 nkivlgmpiygrsf--qntag--igqtyngvgs-gsw--eagiwdykalpk-agatvqyd 335
          :|::|||| |||||   :::   :|    | |: |:   |:|:  |  :    ||| |
Qy    237 SKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRI 296

Db    336 svakgyysynsatkelisfdtpdmintkvaylkslglggsmfw 378
            |  | :    :: ::||  : ::|||:|||  ||||:| |
Qy    297 QDQKVPYIFRD--NQWVGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT    7
ID   CHI1_COCIM     STANDARD;      PRT;   427 AA.
AC   P54196;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   ENDOCHITINASE 1 PRECURSOR (EC 3.2.1.14) (COMPLEMENT-FIXATION ANTIGEN)
DE   (CF-ANTIGEN) (CF-AG).
GN   CTS1.
OS   COCCIDIOIDES IMMITIS.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C735;
RX   MEDLINE; 96144270.
RA   PISHKO E.J., KIRKLAND T.N., COLE G.T.;
RL   GENE 167:173-177(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SILVEIRA;
RA   YANG C., ZHU Y., MAGEE D.M., COX R.A.;
RL   SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; L41663; G1200190; -.
DR   EMBL; U51271; G1256769; -.
DR   EMBL; U33265; G1255728; -.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL;
KW   GLYCOPROTEIN.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?    427       ENDOCHITINASE 1.
FT   CARBOHYD    387    387       POTENTIAL.
FT   CONFLICT     15     47       RWLSLRCLFCELGRFMFTLSTVTAVTVVTDDIQ -> VQAS
FT                                SMSSMPNYYPVPEAPEGGFRSVVYFVNW (IN REF. 2).
FT   CONFLICT    199    199       K -> N (IN REF. 2).
SQ   SEQUENCE   427 AA;  47629 MW;  6CB9AF73 CRC32;

  Query Match           20.1%;  Score 546;  DB 2;  Length 427;
  Best Local Similarity  34.1%;  Pred. No. 2.16e-92;
  Matches    95;  Conservative   71;  Mismatches  94;  Indels  19;  Gaps  17;

Db    115 lkknnrnlktllsiggwty-spnfktpasteegrkkfadtslklmkdlgfdgididweyp 173
          ||| | :|||||:|||| : : :|    :| : |  | :::::::  :|||:|:|||||
Qy     62 LKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDLDWEYP 121

Db    174 ede-kqandfvlllkacrealdaysak-hpngk-kflltiaspagpqny-nk-lklaemd 228
            :   | |   :    ::  :|:    : :|| ::||: | ||| | | :    :  :
Qy    122 GSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIA 180

Db    229 kyldfwnlmaydfsgswdkvsghmsnvfpsttkp-estpfssdkavkdyikagvpankiv 287
            ||| ||||||| |||:||:|| | ::     :  ::::: | || : :  | ||:|::
Qy    181 QNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLI 240

Db    288 lgmplygraf--ast-dg-igtsfngvgg-gsw--engvwdykdmpq-qgaqvteledia 339
          |||| |||:|  ||: |  :|::  | |  |:   | |:  | ::    ||    ::|
Qy    241 LGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQK 300

Db    340 asysydknkrylisydtvkiagkkaeyitkngmgggm-w 377
            :| : ::::  :::| |     |  |:  :|:||:| |
Qy    301 VPYIF-RDNQW-VGFDDVESFKTKVSYLKQKGLGGAMVW 337


RESULT    8
ID   CHIA_SERMA     STANDARD;      PRT;   563 AA.
AC   P07254;
DT   01-APR-1988 (REL. 07, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE A PRECURSOR (EC 3.2.1.14).
GN   CHIA.
OS   SERRATIA MARCESCENS.
```

```
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   ENTEROBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KOO J.C., LIM C.O., CHOI Y.J., KIM C.Y., BAHK J.D., LEE S.Y.,
RA   CHO M.J.;
RL   SUBMITTED (JAN-1993) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 990 / QMB1466;
RA   JONES J.D.G., GRADY K.L., SUSLOW T.V., BEDBROOK J.R.;
RL   EMBO J. 5:467-473(1986).
RN   [3]
RP   REVISIONS, AND X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE; 95219379.
RA   PERRAKIS A., TEWS I., DAUTER Z., OPPENHEIM A.B., CHET I., WILSON K.S.,
RA   VORGIAS C.E.;
RL   STRUCTURE 2:1169-1180(1994).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; L01455; G152818; -.
DR   EMBL; X03657; G46831; -.
DR   PIR; A25090; A25090.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL.
FT   SIGNAL        1     23
FT   CHAIN        24    563       CHITINASE A.
FT   DOMAIN      150    563       CATALYTIC.
FT   ACT_SITE    315    315       PROTON DONOR (PROBABLE).
FT   ACT_SITE    391    391       PROBABLE.
FT   CONFLICT     76     77       TA -> GP (IN REF. 2).
FT   CONFLICT    139    139       A -> P (IN REF. 2).
FT   CONFLICT    226    226       V -> I (IN REF. 2).
FT   CONFLICT    395    395       P -> A (IN REF. 1 AND 2).
FT   CONFLICT    410    430       PAWKPDTAYTTVNGVNALLAQ -> RPGSRHRLHHGERRQC
FT                                AAGQ (IN REF.1 AND 2).
FT   CONFLICT    437    437       V -> I (IN REF. 2).
FT   CONFLICT    464    467       ATGP -> HRA (IN REF. 2).
FT   CONFLICT    473    473       K -> E (IN REF. 2).
FT   CONFLICT    484    484       G -> S (IN REF. 2).
SQ   SEQUENCE   563 AA;  60979 MW;  7764DB57 CRC32;

  Query Match           16.1%;  Score 437;  DB 2;  Length 563;
  Best Local Similarity  28.8%;  Pred. No. 7.14e-68;
  Matches    93;  Conservative   84;  Mismatches 123;  Indels  23;  Gaps  19;

Db    231 pfaalqkaqkgvtawddpykgnfgqlmalkqahpdlkilpsiggwtlsdpfff-mgdkvk 289
          :||:: : | : | |:|     : :: :||  :| || | :|||| ::   |  |    :
Qy     36 AFAGMTNHQLSTTEWNDE-T-LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATAN 93

Db    290 -rdrfvgsvkeflqtwkffdgvdidwefpggkganpnlgspqdgetyvllmkelramldq 348
           |: || |   ||:   | ||:|:|||:||: |: | :    | | :  |: :|   ::|
Qy     94 NRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV----DKERFTTLVQDLANAFQQ 147

Db    349 lsaetgr-kyeltsaisagkdkidkvaynvaq--nsmdhiflmsydfygpfdlknlghqt 405
          :  :|: :  |::|::||   :|  :|:|     ::| : ||:||| |: : |  || :
Qy    148 EAQTSGKERLLLSAAVPAGQTYVD-AGYEVDKIAQNLDFVNLMAYDFHGSWE-KVTGHNS 205

Db    406 alnapawkpdtayt-tvng-vnallaqgvkpgkvvvgtamygrgwtgvngyqnnipftgt 463
          :|      : :| :  |:: |   |  |  ::|:::| : |||: |  :: :  :   :|
Qy    206 PLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPAT 265

Db    464 atg-pvkgtwkngivdyrqiagqfmsgewqytydataeapyvfkpstgdlitfddarsvq 522
          ::| |   |   |:: | :: :  :::  |  :     ||:|:    : : |||  |
Qy    266 GSGTPGPFTKEGGMLAYYEVCS-WKGATKQ-RIQD-QKVPYIFR-DN-QWVGFDDVESFK 320

Db    523 akgkyvldkqlgglfsweidadn 545
          :|  |: :| |||   | :| |:
Qy    321 TKVSYLKQKGLGGAMVWALDLDD 343
```

```
RESULT    9
ID   CHIT_NPVAC     STANDARD;      PRT;   551 AA.
AC   P41684;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   PROBABLE ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
OS   AUTOGRAPHA CALIFORNICA NUCLEAR POLYHEDROSIS VIRUS (ACMNPV).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; BACULOVIRIDAE; EUBACULOVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C6;
RX   MEDLINE; 94303173.
RA   AYRES M.D., HOWARD S.C., KUZIO J., LOPEZ-FERBER M., POSSEE R.D.;
RL   VIROLOGY 202:586-605(1994).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; L22858; G559195; -.
DR   PROSITE; PS00014; ER_TARGET.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    551       PROBABLE ENDOCHITINASE.
FT   ACT_SITE    305    305       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    173    173       POTENTIAL.
FT   CARBOHYD    444    444       POTENTIAL.
FT   SITE        548    551       PREVENT SECRETION FROM ER (POTENTIAL).
SQ   SEQUENCE   551 AA;  61368 MW;  BDB124DB CRC32;

  Query Match           14.5%;  Score 395;  DB 2;  Length 551;
  Best Local Similarity  29.2%;  Pred. No. 1.33e-58;
  Matches    95;  Conservative   80;  Mismatches 122;  Indels  28;  Gaps  27;

Db    221 pwaavqkpqkgvsawn-ep-ykgnfgqlmaaklanphlkilpsiggwtlsdp-fyfmhdv 277
          : |:: : | : : || |  |  :|  |:  |: || || | :|||| ::   |  |
Qy     36 AFAGMTNHQLSTTEWNDETLYQ-EFNGLK--KM-NPKLKTLLAIGGWNFGTQKFTDMVAT 91

Db    278 -ekrnvfvdsvkeflqvwkffdgvdidwefpggkganpslgdadgdaktyillleelram 336
           ::|  ||:|   ||:   | ||:|:|||:||: |: |:: |   : : :  |:::|
Qy     92 ANNRQTFVNSAIRFLRKYSF-DGLDLDWEYPGSQGS-PAV-D-K-E-R-FTTLVQDLANA 144

Db    337 lddleaqt-grv-yeltsaisagydkiavvnya-e-aqkslgkiflmsydfkgawsntdl 392
          ::: |||| |:    |::|::|| :      |  :    :|  : ||:||| |:| :
Qy    145 FQQ-EAQTSGKERLLLSAAVPAG-QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVT- 201

Db    393 gyqttvya-ps-wnseelytthyavdallkqgvdpnkiivgvamygrgwtgvtnytndny 450
          |  : :|      :       ||:  |  |  ::|:|:|:: |||: |  ::
Qy    202 GHNSPLYKRQEESGAAASLNVDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVG 261

Db    451 fsgtgng-pgsgtwedgvvdyrqiqkdlnnyvytfdsaaqasyvfdkskgdlisfdsvds 509
          ::||:| ||: | | |:: | ::          ::     :|:| : : : ::|| |:|
Qy    262 APATGSGTPGPFTKEGGMLAYYEVCSWKGATKQRIQDQ-KVPYIF-RDN-QWVGFDDVES 318

Db    510 vlgkvkyvdrnklgglfaweidadn 534
            || |: :: |||   | :| |:
Qy    319 FKTKVSYLKQKGLGGAMVWALDLDD 343


RESULT   10
ID   CHIA_ALTSO     STANDARD;      PRT;   820 AA.
AC   P32823;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE A PRECURSOR (EC 3.2.1.14) (CHI-A).
GN   CHIA.
OS   ALTEROMONAS SP. (STRAIN O-7).
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   VIBRONACEAE.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93106952.
RA   TSUJIBO H., ORIKOSHI H., TANNO H., FUJIMOTO K., MIYAMOTO K.,
RA   IMADA C., OKAMI Y., INAMORI Y.;
RL   J. BACTERIOL. 175:176-181(1993).
RN   [2]
RP   SEQUENCE OF 22-48, AND CHARACTERIZATION.
RX   MEDLINE; 93099552.
RA   TSUJIBO H., YOSHIDA Y., MIYAMOTO K., IMADA C., OKAMI Y., INAMORI Y.;
RL   CAN. J. MICROBIOL. 38:891-897(1992).
RN   [3]
RP   MUTAGENESIS.
RX   MEDLINE; 93379363.
RA   TSUJIBO H., ORIKOSHI H., IMADA C., OKAMI Y., MIYAMOTO K.,
RA   INAMORI Y.;
RL   BIOSCI. BIOTECHNOL. BIOCHEM. 57:1396-1397(1993).
CC   -!- THE OPTIMUM PH AND TEMPERATURE OF CHI-A ARE 8.0 AND 50 DEGREES
CC       CELSIUS, RESPECTIVELY.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- ENZYME REGULATION: STIMULATED BY MG2+; INHIBITED BY
CC       N-BROMOSUCCINIMIDE AND 2-HYDROXY-5-NITROBENZYL BROMIDE.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; D13762; G216207; -.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL.
FT   SIGNAL        1     21
FT   CHAIN        22    820       CHITINASE A.
FT   ACT_SITE    313    313       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   820 AA;  87346 MW;  838F6EE8 CRC32;

  Query Match            12.7%;  Score 345;  DB 2;  Length 820;
  Best Local Similarity  35.0%;  Pred. No. 9.66e-48;
  Matches    57;  Conservative   44;  Mismatches  47;  Indels  15;  Gaps  13;

Db    250 vysqlmalkqrypdlkilpsvggwtlsdp-fhgftnkan-rdtfvasvkqflktwkfydg 307
          :| :: :||   | || | ::||| ::   |      || |:||| |  :||:   | ||
Qy     55 LYQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYSF-DG 113

Db    308 vdidwefpggdgpnpdlgdpindgpayvalmqelramldeleaet-grq-yeltsaigag 365
          :|:|||:||::|: | : |  ::   : :|:|:|   ::: ||:| |::   |::|: ||
Qy    114 LDLDWEYPGSQGS-PAV-D--KE-R-FTTLVQDLANAFQQ-EAQTSGKERLLLSAAVPAG 166

Db    366 ydkiedvdyqaaq--qymdyifamtydfygawnnetghqtgiy 406
            :   |  |:     | :|::  |:||| |:|:: ||| : :|
Qy    167 -QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLY 208


RESULT   11
ID   CHIB_SERMA     STANDARD;      PRT;   499 AA.
AC   P11797;
DT   01-OCT-1989 (REL. 12, CREATED)
DT   01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE B PRECURSOR (EC 3.2.1.14).
GN   CHIB.
OS   SERRATIA MARCESCENS.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   ENTEROBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 990 / QMB1466;
RX   MEDLINE; 89345110.
RA   HARPSTER M.H., DUNSMUIR P.;
RL   NUCLEIC ACIDS RES. 17:5395-5395(1989).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; X15208; G47228; -.
DR   PIR; S04856; S04856.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL.
FT   SIGNAL        1     41
FT   CHAIN        42    499       CHITINASE B.
FT   ACT_SITE    144    144       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   499 AA;  55464 MW;  9A585FFC CRC32;

  Query Match            12.4%;  Score 337;  DB 2;  Length 499;
  Best Local Similarity  34.0%;  Pred. No. 5.03e-46;
  Matches    54;  Conservative   47;  Mismatches  44;  Indels  14;  Gaps  10;

Db     78 ltalkahnpslrimfsiggwyysndlgvshanyvnavktpaartkfaqscvrimkdygfd 137
          : :||  || |: :::|||||   | :| :: :: : | |:  |  |  | :|::: |:||
Qy     59 FNGLKKMNPKLKTLLAIGGW---N-FG-TQ-KFTDMVATANNRQTFVNSAIRFLRKYSFD 112

Db    138 gvdidweyp--qaa-evd-g-fiaalqeirtllnqqtiadgrqalpyqltiagaggaffl 192
          |:|:|||||  |::  ||   | : :|::   : |:: : |:: |   |: | ::|  ::
Qy    113 GLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERL--LLSAAVPAGQTYV 170

Db    193 sryysklaqivapldyinlmtydlagpwekitnhqaalf 231
              :  |   ||::|||:||: |:||||:| | ::|:
Qy    171 DAGYE-VDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLY 208


RESULT   12
ID   CHIT_STRLI     STANDARD;      PRT;   619 AA.
AC   P36909;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE C PRECURSOR (EC 3.2.1.14).
GN   CHIC.
OS   STREPTOMYCES LIVIDANS.
OC   PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93294525.
RA   FUJII T., MIYASHITA K.;
RL   J. GEN. MICROBIOL. 139:677-686(1993).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- INDUCTION: BY CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS A BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC       (CBD).
DR   EMBL; D12647; G391910; -.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS00018; EF_HAND.
DR   PROSITE; PS00561; CBD_BACTERIAL.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    619       CHITINASE C.
FT   DOMAIN       35    140       CELLULOSE-BINDING.
FT   DOMAIN      148    230       FIBRONECTIN TYPE-III.
FT   DOMAIN      240    619       CATALYTIC.
FT   ACT_SITE    382    382       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   619 AA;  65200 MW;  23E3E1EC CRC32;

  Query Match            11.1%;  Score 301;  DB 2;  Length 619;
  Best Local Similarity  28.2%;  Pred. No. 2.19e-38;
  Matches    86;  Conservative   65;  Mismatches 129;  Indels  25;  Gaps  18;

Db    311 w-dqplrgnfnqlrklkakypnikilysfggwtwsgg-fpdavknpaafakschdlvedp 368
          | |: |  :|| |:|::   |::| | ::|||  :   | | |  :       :  :
Qy     50 WNDETLYQEFNGLKKMN---PKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFL 106

Db    369 rwadvfdgidldweypnacglscdetsapnafssmmkamraefgqdylita-a---vtad 424
          |    |||:||||||| : | |   :   : |:::: :   | |:   ::     ::|
Qy    107 RKYS-FDGLDLDWEYPGSQG-S--PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAA 162
```

```
Db    425 gsdggk-idaa-dygeaskyidwynvmtydffgawak-ngptaphspltaydgipqqgfn 481
          : |   :||: :    :  :|  |:|:||| |:| |  |  :|       | :  ::|
Qy    163 VPAGQTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA-SLN 221

Db    482 tadamakfkskgvpadklligigfygrgwt-gvtqsap-ggtatgpat-gty--eagied 536
            |:     || || ||::|:  |||: | : :  :  |: |||::| | :  |:|:
Qy    222 VDAAVQQWLQKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLA 281

Db    537 y-kvlknscpatgtiag--tayahcgsnwwsydtpatikskmdwaeqqglggaffwefsg 593
          | |     ::   |     :|    : | ::|   ::|:|:    | |||||  | :
Qy    282 YYEVCSWKGATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDL 341

Db    594 dtang 598
          |  |
Qy    342 DDFAG 346


RESULT   13
ID   CHIT_STRPL     STANDARD;      PRT;   610 AA.
AC   P11220;
DT   01-JUL-1989 (REL. 11, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CHITINASE 63 PRECURSOR (EC 3.2.1.14).
GN   CHTA.
OS   STREPTOMYCES PLICATUS.
OC   PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92192480.
RA   ROBBINS P.W., OVERBYE K., ALBRIGHT C., BENFIELD B., PERO J.;
RL   GENE 111:69-76(1992).
RN   [2]
RP   SEQUENCE OF 1-45 FROM N.A., AND SEQUENCE OF 31-45.
RX   MEDLINE; 88087127.
RA   ROBBINS P.W., ALBRIGHT C., BENFIELD B.;
RL   J. BIOL. CHEM. 263:443-447(1988).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- INDUCTION: BY CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS A BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC       (CBD).
DR   EMBL; M82804; G153216; -.
DR   EMBL; M18397; G153209; -.
DR   PIR; A29912; A29912.
DR   PIR; JH0573; JH0573.
DR   HSSP; P07254; 1CTN.
DR   PROSITE; PS00018; EF_HAND.
DR   PROSITE; PS00561; CBD_BACTERIAL.
DR   PROSITE; PS01095; CHITINASE_18.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL.
FT   SIGNAL        1     30
FT   CHAIN        31    610       CHITINASE 63.
FT   DOMAIN       35    140       CELLULOSE-BINDING.
FT   DOMAIN      148    229       FIBRONECTIN TYPE-III.
FT   DOMAIN      236    610       CATALYTIC.
FT   ACT_SITE    383    383       PROTON DONOR (BY SIMILARITY).
FT   CONFLICT      3      3       F -> I (IN REF. 2).
SQ   SEQUENCE   610 AA;  63974 MW;  2F5E8E35 CRC32;

  Query Match           10.6%;  Score 287;  DB 2;  Length 610;
  Best Local Similarity  26.4%;  Pred. No. 1.86e-35;
  Matches    78;  Conservative   63;  Mismatches 134;  Indels  21;  Gaps  16;

Db    320 fnqlrnlkaeyphikilysfggwtwsgg-fpdavknpaafakschdlvedprwadvfdgi 378
          : ::  ||   | :| | ::|||  :   | | |  :       :  :   |    |||:
Qy     56 YQEFNGLKKMNPKLKTLLAIGGWNFGTQKFTDMVATANNRQTFVNSAIRFLRKYS-FDGL 114

Db    379 dldweypnacglscdetsapnafssmmkamraefgqdylita-a---vtadgsdggk-id 433
          |||||| : | |   :   : |::::  :   | |:   ::      ::|  : |   :|
Qy    115 DLDWEYPGSQG-S--PAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAGQTYVD 171

Db    434 aa-dygeaskyidwynvmtydffgawak-ngptaphsplnaydgipqqgfttadamakfk 491
          |: :    :  :|  |:|:||| |:| |  |  :|       | :  ::    |:
Qy    172 AGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAA-SLNVDAAVQQWL 230

Db    492 skgvpadklligigfygrgwt-gvtqsap-ggtatgpaa-gty--eagiedy-kvlknsc 545
          || || ||::|:  |||: | : :  :  |: |||::: | :  |:|:  |  |
Qy    231 QKGTPASKLILGMPTYGRSFTLASSSDTRVGAPATGSGTPGPFTKEGGMLAYYEVCSWKG 290

Db    546 patgtvag--tayahcgtnwwsydtpatikskmdwaeqqglggaffwefsgdttng 599
          ::   :     :|      | ::|   ::|:|:    | |||||  | :  |   |
Qy    291 ATKQRIQDQKVPYIFRDNQWVGFDDVESFKTKVSYLKQKGLGGAMVWALDLDDFAG 346


RESULT   14
ID   KTXA_KLULA     STANDARD;      PRT;  1146 AA.
AC   P09805;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   KILLER TOXIN ALPHA AND BETA SUBUNITS PRECURSOR (RF2 PROTEIN)
DE   (ENDOCHITINASE (EC 3.2.1.14)).
OS   KLUYVEROMYCES LACTIS (YEAST).
OG   PLASMID PGKL-1.
OC   EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS 2359/152;
RA   SOR F., FUKUHARA H.;
RL   CURR. GENET. 9:147-155(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84297209.
RA   STARK M.J.R., MILEHAM A.J., ROMANOS M.A., BOYD A.;
RL   NUCLEIC ACIDS RES. 12:6011-6030(1984).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85037931.
RA   TOKUNAGA M., KAWAMURA A., HISHINUMA F.;
RL   NUCLEIC ACIDS RES. 12:7581-7597(1984).
RN   [4]
RP   IDENTIFICATION OF PROTEIN, AND SEQUENCE OF 30-44 AND 895-916.
RX   MEDLINE; 87004569.
RA   STARK M.J.R., BOYD A.;
RL   EMBO J. 5:1995-2002(1986).
RN   [5]
RP   SIMILARITY TO CHITINASE OF ALPHA-SUBUNIT.
RX   MEDLINE; 90259069.
RA   BRADSHAW H.D. JR.;
RL   NATURE 345:299-299(1990).
RN   [6]
RP   CHITINASE ACTIVITY OF ALPHA-SUBUNIT.
RX   MEDLINE; 91301161.
RA   BUTLER A.R., O'DONNEL R.W., MARTIN V.J., GOODAY G.W., STARK M.J.R.;
RL   EUR. J. BIOCHEM. 199:483-488(1991).
CC   -!- FUNCTION: THE ALPHA SUBUNIT IS A POTENT EXOCHITINASE. ALONG WITH
CC       THE BETA SUBUNIT IT PLAYS A ROLE IN THE INITIAL INTERACTION OF THE
CC       TOXIN WITH SENSITIVE CELLS AND ALLOW THE GAMMA SUBUNIT (THE ACTIVE
CC       TOXIN) TO GAIN ENTRY INTO THE CELL.
CC   -!- PTM: RF2 IS POTENTIALLY SPLIT BY MEMBRANE-BOUND BASIC AMINO ACID-
CC       SPECIFIC PEPTIDASE TO YIELD THE ALPHA AND BETA SUBUNITS.
CC   -!- SUBUNIT: THE KILLER TOXIN IS COMPOSED OF THREE SUBUNITS: ALPHA,
CC       BETA AND GAMMA.
CC   -!- SIMILARITY: LOCAL, TO OTHER CHITIN-BINDING PROTEINS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
DR   EMBL; X07127; G2829; -.
DR   EMBL; X00762; G2844; -.
DR   EMBL; X01095; G2849; -.
DR   PIR; S07915; S07915.
DR   HSSP; P02877; 1HEV.
DR   PROSITE; PS00026; CHITIN_BINDING.
```

```
DR   PROSITE; PS01095; CHITINASE_18.
KW   PLASMID; TOXIN; GLYCOPROTEIN; SIGNAL; HYDROLASE; GLYCOSIDASE;
KW   CHITIN DEGRADATION.
FT   SIGNAL        1     17       POTENTIAL.
FT   PROPEP       18     29       POTENTIAL.
FT   CHAIN        30    892       ALPHA SUBUNIT (PROBABLE).
FT   CHAIN       895   1146       BETA SUBUNIT (PROBABLE).
FT   DOMAIN      310    363       CHITIN-BINDING (BY SIMILARITY).
FT   DOMAIN      364    635       CATALYTIC.
FT   ACT_SITE    495    495       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    771    771       POTENTIAL.
FT   CARBOHYD    858    858       POTENTIAL.
FT   CARBOHYD    868    868       POTENTIAL.
FT   CARBOHYD    876    876       POTENTIAL.
FT   CARBOHYD    995    995       POTENTIAL.
FT   CARBOHYD   1082   1082       POTENTIAL.
FT   CARBOHYD   1117   1117       POTENTIAL.
SQ   SEQUENCE   1146 AA;  128936 MW;  CDD1620F CRC32;

  Query Match             9.7%;  Score 263;  DB 5;  Length 1146;
  Best Local Similarity  28.4%;  Pred. No. 1.67e-30;
  Matches    60;  Conservative   58;  Mismatches  74;  Indels  19;  Gaps  14;

Db    425 ddsafsksaflkvtssk-k-ipsfggwdfstspstytifrnavktdqnrntfannlinfm 482
          |:: :     ||  ::| | : ::|||:|:|    :|   : | |  || || |: | |:
Qy     52 DETLYQEFNGLKKMNPKLKTLLAIGGWNFGT-Q-KFT---DMVATANNRQTFVNSAIRFL 106

Db    483 nkynldgidldweypgapdipdipaddsssgsnyltflkllkgkmps-gktl-siaipss 540
           ||::||:||||||||:   | :  :  ::    |:      :  ::  : | | |:|::
Qy    107 RKYSFDGLDLDWEYPGSQGSPAVDKERFTTLVQDLANAFQQEAQTSGKERLLLSAAVPAG 166

Db    541 ywylkn-fpisdiqntvdymvymtydihgiwe--ygkans-yi---nchtprk-eiedai 592
            |:   : :  |   :|::  |:||:|| ||   |  :: |    :  ::   ::: |:
Qy    167 QTYVDAGYEVDKIAQNLDFVNLMAYDFHGSWEKVTGHNSPLYKRQEESGAAASLNVDAAV 226

Db    593 km-ldkagvkfnkvfggvanygrsykmvntn 622
             |:| |   :|:: |:: |||| : :::
Qy    227 QQWLQK-GTPASKLILGMPTYGRSFTLASSS 256


RESULT   15
ID   GP39_BOVIN     STANDARD;      PRT;    36 AA.
AC   P30922;
DT   01-JUL-1993 (REL. 26, CREATED)
DT   01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   CARTILAGE GLYCOPROTEIN-39 (GP-39) (39 KD WHEY PROTEIN) (FRAGMENT).
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; ARTIODACTYLA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 88106603.
RA   REJMAN J.J., HURLEY W.L.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 150:329-334(1988).
CC   -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE CAPACITY OF CELLS TO
CC       RESPOND TO AND COPE WITH CHANGES IN THEIR ENVIRONMENT.
CC   -!- PTM: GLYCOSYLATED.
CC   -!- TISSUE SPECIFICITY: MAMMARY SECRETIONS COLLECTED DURING THE
CC       NONLACTATING PERIOD.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
DR   PIR; A27682; A27682.
KW   GLYCOPROTEIN.
FT   NON_TER      36     36
SQ   SEQUENCE   36 AA;  4264 MW;  F3935740 CRC32;

  Query Match             6.3%;  Score 172;  DB 4;  Length 36;
  Best Local Similarity  48.6%;  Pred. No. 8.53e-13;
  Matches    17;  Conservative   11;  Mismatches   7;  Indels   0;  Gaps   0;

Db      2 klixyytswsqyregdgsxfpdaidpflxthviys 36
          || |:|:|:|||:|::  :|  :|| | ||:||:
```

```
Qy      2 KLVCYFTNWAQYRQGEARFLPKDLDPSLCTHLIYA 36


Search completed: Fri Jun 27 17:08:13 1997
Job time : 60 secs.
```